3. COMPARISONS AMONG EXPOSURE GROUPS

3.1 Allocation of person-years to time-dependent exposure categories
3.2 Grouped data from the Montana smelter workers study
3.3 Comparison of directly standardized rates
3.4 Comparison of standardized mortality ratios
3.5 Comparison of internally standardized mortality ratios
3.6 Preferred methods of analysis of grouped data
3.7 Proportional mortality and dose-response analyses

CHAPTER 3

COMPARISONS AMONG EXPOSURE GROUPS

The techniques of standardization introduced in the last chapter are typically used to determine whether the cause-specific mortality rates for the study cohort are comparable with those from some appropriate standard population. The observation of an elevated CMF or SMR for a particular cause of death may alert the investigator to the possibility that the cohort members are subject to exposures which increase their risk for that disease. However, a single elevated mortality ratio is usually not regarded in itself as sufficient evidence for a causal relationship, unless it is extremely large. A much better indication of causality is the demonstration of a trend in the mortality ratios with degree or duration of exposure.

In this chapter, we explore several elementary methods used by epidemiologists and biostatisticians to examine cohort data for evidence of differences in death rates between subgroups defined by exposures or other factors, and in particular for evidence of dose-response relationships. The most appropriate methods are adaptations of the classical Mantel–Haenszel analyses of grouped case-control data presented in Chapter 4 of Volume 1. These are covered in §3.6 below. Earlier sections consider methods based on the standardization techniques developed in the last chapter. These are of interest largely for historical reasons. Both the limitations and the potential of the various techniques are illustrated by their systematic application to the Montana smelter workers study. In addition, we cite several examples from the literature which point up notable innovations or pitfalls in the use of these statistical tools.

3.1 Allocation of person-years to time-dependent exposure categories

The first step in comparing death rates among different subgroups of the cohort is simply to estimate the rates for each of them separately using the techniques outlined in the previous chapter. This is quite straightforward when the subgroups are formed on the basis of information available at entry into the study – for example, when they are defined by age or calendar period at entry or by a classification of the initial work area according to measured levels of exposure. One simply treats each subgroup as a separate cohort and carries out the usual allocation of deaths and calculation of person-years by age and time for each one independently. Since a study member contributes person-years observation to only one subgroup, there is no ambiguity about the assignment of his observation time.

It is also of interest, however, to make comparisons among subgroups defined on the basis of variables that change values as the subject moves through the study. For example, subjects often continue to accumulate exposures of interest during the same period that they are being followed for evaluation of cause-specific mortality. Industrial workers may be entered on study while still relatively young and be followed through their working years and on into retirement. If the measured exposures are distributed continuously over the working lifetime, the subjects with the highest cumulative levels of exposure are frequently those who have lived the longest. This is even truer when a variable that reflects duration of exposure is being analysed for its relationship to the risk of disease. Special precautions are required to ensure that the allocation of person-years is made appropriately.

Several investigators have attempted to establish a dose-response trend in such circumstances by classifying each subject into a single subgroup on the basis of his total cumulative exposure or duration of employment at the end of the study. Mortality ratios computed separately for each subgroup are then compared. Unfortunately, results obtained in this manner are fallacious, since the early person-years of follow-up, when cumulative exposures are light, are being allocated to the same heavy exposure category as the later person-years. The death rates calculated in this fashion for the highest exposure categories are too low, since person-years during which no death could have occurred are included in the denominator. Rates for the lowest exposure categories are too high since it is only the individuals who die with short exposures who contribute to the denominator; the person-years of someone who might have died with short-term exposure, but in fact did not, are allocated elsewhere.

The correct assignment of each increment in person-years of follow-up is to that same exposure category to which a death would be assigned should it occur at that time. Subjects who change their exposure classification as they move through the study, as many in fact do, thus contribute to the person-years denominators of the rates for several exposure categories. Figure 3.1 illustrates schematically the proper, dynamic method of allocation as well as the improper, fixed method when duration of follow-up itself is used to define the subgroups being compared.

Table 3.1 presents an example of the magnitude of this dose-response fallacy in actual practice. In the original report of an early study of vinyl chloride workers (Duck *et al.*, 1975), the authors observed that the all-causes SMR declined from 110 for those employed for less than 15 years to 61 for those employed for 15 or more years and stated that no significant excess of mortality had occurred. However, the apparent decline in the SMRs was due entirely to the use of an improper methodology. After correcting the fixed person-years allocation used in the original analysis to an appropriate, dynamic one, the statistically significant negative trend in the SMRs disappeared. There was even an indication of a positive trend in the SMR for digestive cancer with duration of exposure (Duck & Carter, 1976; Wagoner *et al.*, 1976). Enterline (1976) discusses a similar error in the report of Mancuso and El-Attar (1967), who failed to detect a trend in respiratory cancer SMRs among asbestos workers who had been employed for increasing lengths of time.

We describe two algorithms for the correct assignment of person-years observation in the presence of time-dependent exposures categories, the use of which enables one

Fig. 3.1 Schematic diagram illustrating proper and improper methods of allocation of person–years. ×, death from cause of interest; ○, withdrawal

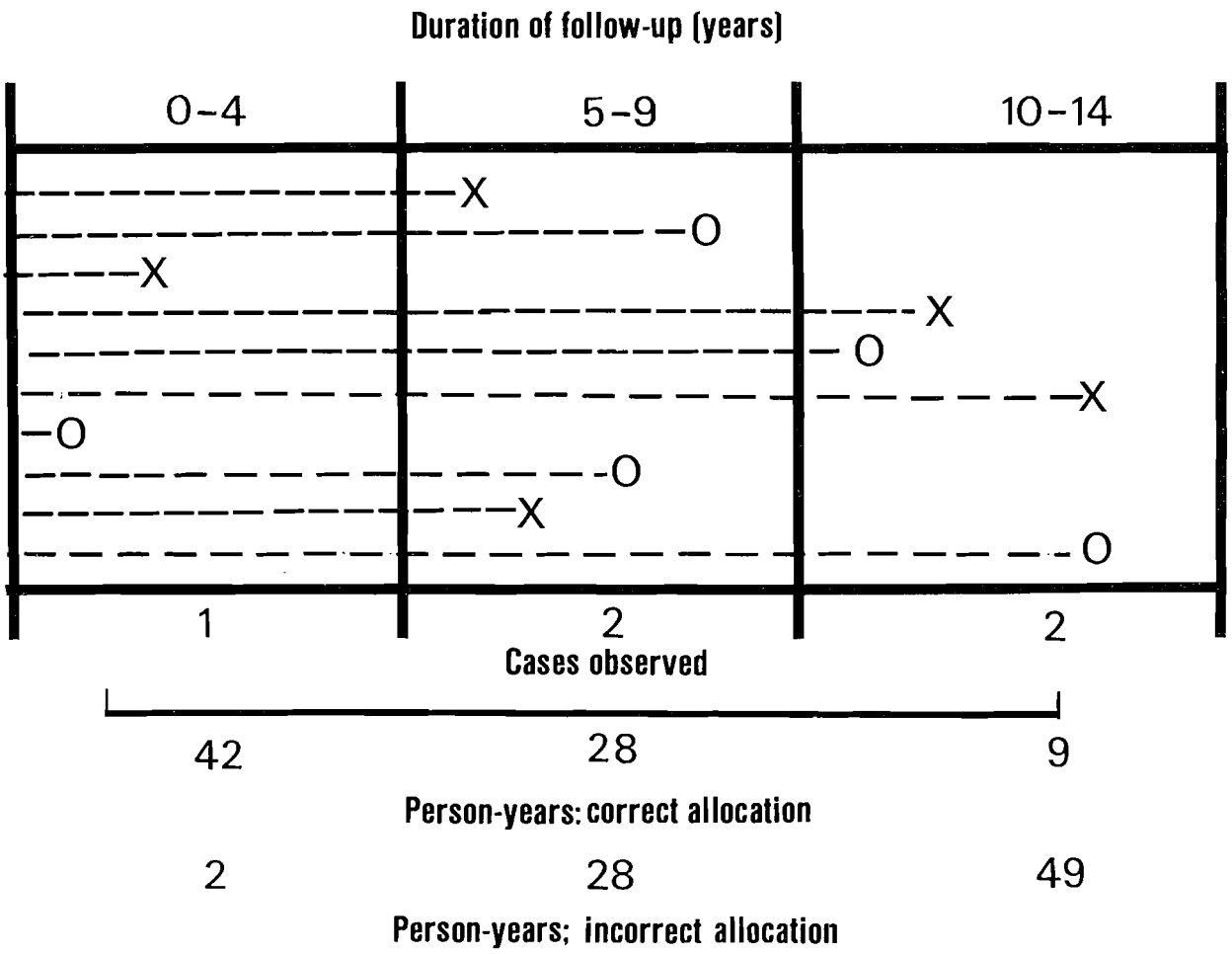


Table 3.1 Reanalysis of data by Duck *et al.* showing original *versus* revised numbers of expected deaths and SMRs by duration of exposure and cause of death[a]

Cause of death	Duration of exposure (years)	No. of observed deaths	No. of expected deaths		SMR	
			Original	Revised	Original	Revised
All causes	0–14	111	100.92	118.97	110	94
	15+	25	41.30	24.15	61	104
Total cancers	0–14	27	25.55	29.93	106	90
	15+	8	10.89	6.51	73	123
Digestive system cancers	0–14	7	7.77	9.10	90	77
	15+	4	3.31	1.98	121	202
Lung cancer	0–14	13	10.73	12.57	121	103
	15+	3	4.80	2.96	62	101

[a] From Duck *et al.* (1975); Duck & Carter (1976)

to avoid the dose-response fallacy caused by the overlapping of exposure and follow-up periods.

(*a*) *Algorithms for exact allocation of person-years*

In practice, there may be several time-dependent exposure variables of interest, and a simultaneous classification of deaths and person-years in a multidimensional table is required. For example, in addition to duration of time since start of employment, we usually need to keep track of age and calendar year, if only for purposes of standardization. Time since cessation of exposure adds a fourth dimension. Determination of the exact length of observation time that each individual contributes to each cell in the four-way table may seem initially to present a difficult problem.

Clayton (1982) describes a computing algorithm for making the appropriate allocation of person-years in such circumstances. It requires that one have available exact dates of entry into and exit from the various time-dependent classes. While not the most efficient method for all problems, this procedure has the advantage of simplicity and generality. Suppose, for example, that one wishes to determine the person-years observation time contributed by one subject to the cell defined by the age range 40–49, the calendar period from 1950–1954, and the interval from five to ten years since first exposure to some risk factor. Then Clayton's procedure is as follows:

(A) Choose the *latest* of the three dates: date of birth +40 years, 31 December 1949, and date of first exposure +five years.

(B) Choose the *earliest* of the four dates: date of birth +50 years, 31 December 1954, date of first exposure +ten years, and date of exit from study.

(C) If B precedes A, then the individual makes no contribution to this cell. Otherwise, the observation time contributed is the time interval from date A to date B.

The calculation must be repeated for each individual for each such cell in the multidimensional table (three dimensions in this example). It accommodates time-dependent variables defined in terms of cumulative length of exposure to particular agents, provided that one knows the exact dates at which cell boundaries are crossed. For example, one could add to the above specifications the requirement that the individual has received a cumulative exposure of between 5 and 10 units of radiation while employed in a nuclear industry. If periodic readings of radiation exposure were made, so that the dates of crossing the 5 and 10 unit boundaries could be estimated, these two dates would be added to those in parts (A) and (B) above.

An alternative, more efficient algorithm (Clayton, personal communication) is available when all of the axes of the multidimensional classification represent time variables that advance in pace with one another (age, calendar year, duration of time from initial exposure) rather than variables such as cumulative exposure or duration of (intermittent) employment, which advance at varying rates depending upon the entire history. This algorithm is presented in Appendix IV.

When using one of the standard programmes for cohort analysis it may be feasible to obtain the number of deaths and person-years in each age-time-exposure category by making separate passes through the data for each exposure category. One defines the

dates of entry into and exit from the 'study' for each individual to correspond to the dates of entry into and exit from the particular exposure category. However, this approach is too cumbersome and inefficient to be practical when the number of separate exposure categories is very large.

(*b*) *Approximate methods of allocating person-years*

A drawback to Clayton's algorithms is that they require the exact dates at which an individual crosses from each time-dependent cell into another. In practice, exact dates may be available for some of the relevant variables but not others. For example, we may know a worker's birthdate and date of termination, but have available only the (integral) age and calendar year at which he entered the study or moved between jobs. Nevertheless, it may be possible in such cases to assign approximate dates to the relevant events so that a consistent ordering is maintained between the dates of entry, first exposure, termination and so on, and Clayton's method may then be used. It is important that the same procedure be applied also to the classification of deaths, so that one does not have person-years accumulating in cells where no deaths are possible, or *vice versa.*

An alternative approach to the problem of missing days and months in date variables is to use an approximate method of person-years allocation based on integral ages and calendar years. One such method was outlined in §2.1. For some of the examples in this monograph we have employed yet another approximation which divides each subject's observation period into annual intervals that are allocated in their entirety to a given time/exposure cell. Specifically, at the midpoint of each calendar year of follow-up, a determination is made as to the cell in which the subject should be classified at that moment. All of the observation time for that year, which may be less than a full year in case of entry into or exit from the study, is allocated to the one cell.

3.2 Grouped data from the Montana smelter workers study

One of the major themes of this monograph is the statistical analysis of grouped cohort data consisting of cause-specific deaths and person-years denominators classified by age, calendar period and relevant exposure variables, some of which may be time-dependent. In order to illustrate and compare the various analytical approaches, and to provide the reader with material that he can use to test his comprehension of the methodology, it is helpful to have available a data set that is reasonably typical of what one encounters in practice. Of course, one needs to balance the realism of the example against the need for simplicity if it is to be used as a pedagogic device.

For this purpose we used the approximate method of person-years allocation just mentioned to summarize the data from the Montana study into a three-way table with the dimensions age, calendar period and arsenic exposure. Cumulative exposure was measured in terms of the duration of time spent in certain areas of the smelter where airborne arsenic levels were thought to be higher than average. It thus represents a relatively crude way of separating workers (or, more precisely, their person-years of observation) according to the presumed degree of hazard. The largely descriptive

analyses consist of estimating separate summary mortality measures for each exposure category and testing the statistical significance of the differences, especially for evidence of a trend with increasing exposure. Other, more refined approaches to dose-time-response analyses are discussed in Chapter 6.

The epidemiologists who conducted this study classified the 30 work areas within the plant into three levels of arsenic exposure (see Appendix IE). 'High' arsenic exposure areas comprised the arsenic kitchen, arsenic roaster and cottrell, whereas those with 'moderate' exposure levels were the convertor, reverbatory furnace, ore roaster and acid plant. All other areas were regarded as giving only 'light' exposures (Lee & Fraumeni, 1969). From the original data file containing the dates of entry into and exit from each work area for each worker, summary data consisting of the number of years worked in both high and moderate exposure areas were recorded by five-year calendar periods starting in 1910. By assuming that the exposure intensity was constant during each such period, we were able to determine the appropriate exposure duration category into which each individual should be classified at each point in time: (i) under 1.0 years moderate or high arsenic exposure; (ii) 1.0–4.9 years; (iii) 5.0–14.9 years; and (iv) 15 or more years.

The assignment of an exposure category to each calendar year was based on the duration of heavy/medium exposure experienced at a point two years earlier. Such adjustments to cumulative exposure variables are a crude way of coping with the bias that can arise from the fact that workers who have just entered a new cumulative exposure category are necessarily still employed and thus at lower risk of death, whereas those who change employment or retire for health reasons may have higher death rates (Gilbert, 1983). See the discussion in §1.5*c* of the selection biases, known collectively as the 'healthy worker effect', that are caused by the fact that health status has a major influence on hiring, job changes and termination. This adjustment would be less necessary if it were possible to use onset of disease as the endpoint, rather than death from disease, since onset presumably occurs closer to the time of any adverse health effect.

For the descriptive analyses reported in this chapter, the cohort was divided into two subcohorts, one consisting of the 1482 men employed prior to 1925 and the other of the remainder. The reason for this division was the fact that the selective flotation process introduced in 1924 apparently resulted in greatly reduced arsenic exposures (Lee-Feldstein, 1983). Substantially different dose-response trends are evident in the two groups. An alternative and possibly more appropriate means of coping with the change would be to classify the exposures as to the period during which they were actually received, namely before or after 1 January 1925. A man hired prior to 1925 could contribute to both sets of exposure duration variables, while someone employed later would contribute only to the post-1925 categories. However, this refinement is too complex for illustrative purposes.

We used four ten-year age groups of 40–49, 50–59, 60–69 and 70–79 years and four calendar periods, 1938–1949, 1950–1959, 1960–1969 and 1970–1977, in order to keep the data file to a reasonable size. In actual practice, five-year intervals of age by calendar year (quinquinquennia) would be considered more appropriate in order to take full account of their potentially confounding effects. In order to be able to

Table 3.2 Standard respiratory cancer death rates and standard weights used for comparative analyses of the Montana smelter workers data

Age range (years)	No. of deaths per 1000 person-years Calendar period				Standard weight (%)
	1938–1949	1950–1959	1960–1969	1970–1977	
40–49	0.14817	0.21896	0.28674	0.37391	37.4
50–59	0.47412	0.80277	1.05824	1.25469	30.1
60–69	0.73136	1.55946	2.33029	2.90461	21.5
70–79	0.73207	1.63585	2.85724	4.22945	11.0

calculate and compare SMRs for the various exposure classes, standard respiratory cancer death rates were determined for each of the 16 age/calendar cells by taking a weighted average of the death rates for the corresponding quinquinquennia (Appendix III), using weights proportional to the observed person–years. For calculation of directly standardized rates by exposure class, we chose weights to be proportional to the age distribution of the 1950 US population (Table 2.5). These weights thus depend only on age and not on calendar year. The standard rates and weights are both shown in Table 3.2.

Table 3.3 presents summary data on the numbers of respiratory cancer deaths and person-years allocated to each exposure category by this method, as well as the results of certain analyses described below. Deaths and person-years that occurred outside the age range 40–79 years are ignored. The entire set of data records, consisting of observed respiratory cancer deaths and person-years denominators for each combination of age, period and exposure, as well as other data, is listed in Appendix V. Note that age-year-exposure categories with no person-years of observation are omitted. The omissions are due largely to the fact that persons hired before 1925 could not contribute observations to the younger age groups during the later calendar intervals.

A major weakness of the Lee and Fraumeni study, which also affects all the analyses of the Montana data reported in this monograph, is the lack of smoking histories for the 8014 smelter workers. Welsh *et al.* (1982) subsequently ascertained smoking information by mail questionnaire or telephone interviews from a random sample of 1800 men, using proxy respondents for men who had died. They reported that the percentage of smokers was higher than for the USA as a whole, and this could well explain the high rates of respiratory cancer and ischaemic heart disease even in the 'low' exposure category. There was little difference in smoking habits among men in the arsenic categories, however, so that the dose-response relationships are unlikely to be confounded by smoking. However, the positive effects of certain other variables on respiratory cancer, notably foreign birthplace, could well be secondary to the effects of smoking. Unfortunately, the smoking data were not available and could not be considered in the illustrative analyses.

Table 3.3 Dose-response analysis of respiratory cancer deaths among Montana smelter workers, based on external standardization

	Cumulative years of moderate/heavy arsenic exposure (lagged two years)				
	0–0.9	1.0–4.9	5.0–14.9	15+	Total
Workers employed prior to 1925					
No. of observed deaths	51	17	13	34	115
Person-years (×1000)	19.017	2.683	2.600	3.871	28.171
Crude rate (per 1000 person-years)	2.681	6.337	5.000	8.783	4.082
Standardized rate (per 1000 person-years)	2.641	7.433	5.832	7.397	4.215
Standard population rate[a] (per 1000 person-years)	1.185	1.185	1.185	1.185	1.185
Expected deaths (E_k^*) (standard population)	21.47	2.95	2.76	4.44	31.62
CMF (%)	222.8	627.0	492.0	624.0	355.6
SMR (%)	237.5	577.1	471.7	765.0	363.7
Relative risk (ratio of SMRs)	1.0	2.43	1.99	3.22	
Adjusted expected ($\tilde{E}_k^*$)	78.10	10.71	10.02	16.17	115.00
Test for homogeneity of SMR: $\chi_3^2 = 33.7$; test for trend: $\chi_1^2 = 30.5$					
Workers employed 1925 or later					
No. of observed deaths	100	38	15	8	161
Person-years (× 1000)	74.677	13.693	5.940	2.510	96.820
Crude rate (per 1000 person-years)	1.339	2.775	2.525	3.187	1.663
Standardized rate (per 1000 person-years)	1.557	2.409	2.482	3.949[b]	1.824
Standard population rate[a] (per 1000 person-years)	1.031	1.031	1.031	1.057[b]	1.031
Expected deaths (E_k^*) (standard population)	74.12	13.84	6.83	3.66	98.46
CMF (%)	155.1	233.7	240.8	373.6[b]	177.0
SMR (%)	134.9	274.6	219.6	218.4	163.5
Relative risk (ratio of SMRs)	1.0	2.04	1.63	1.62	
Adjusted expected ($\tilde{E}_k^*$)	121.21	22.63	11.17	5.99	161.00
Test for homogeneity of SMR: $\chi_3^2 = 16.14$; test for trend: $\chi_1^2 = 8.74$					

[a] See Example 3.1

[b] Based on 14 age × calendar periods for which data are available (see Appendix V) and therefore not comparable to the others. The other exposure categories have data for all 16 age × calendar periods.

3.3 Comparison of directly standardized rates

The goal of a comparative analysis is to describe the effects of the different levels of exposure on death rates from particular diseases. Ideally, this should be done at fixed levels of potentially confounding variables such as age and calendar year. However, the large number of comparisons and the instability of the component rates would then

Table 3.4 Notation used for two-way classification of deaths and person-years

Stratum (j)		Exposure level (k)				
		1	2	...	K	Total
1	Deaths	d_{11}	d_{12}	...	d_{1K}	D_1
	Person-years	n_{11}	n_{12}	...	n_{1K}	N_1
2	Deaths	d_{21}	d_{22}	...	d_{2K}	D_2
	Person-years	n_{21}	n_{22}	...	n_{2K}	N_2
⋮		⋮	⋮	⋮	⋮	⋮
J	Deaths	d_{J1}	d_{J2}	...	d_{JK}	D_J
	Person-years	n_{J1}	n_{J2}	...	n_{JK}	N_J
Total	Deaths	O_1	O_2	...	O_K	O_+
	Person-years	n_{+1}	n_{+2}	...	n_{+K}	$N_+ = n_{++}$

make for a rather confusing picture. In our example, 16 separate evaluations depending on the particular age/year stratum would be required. One possible remedy is to base the evaluation on a summary measure such as the directly standardized rate.

Table 3.4 introduces some notation for the number of deaths and person-years of observation in each of J strata ($j = 1, \ldots, J$) and K exposure categories ($k = 1, \ldots, K$). Thus, the directly standardized rate for the kth exposure level may be written

$$\hat{\Lambda}_k = \sum_{j=1}^{J} w_j d_{jk}/n_{jk}, \tag{3.1}$$

where the weights are assumed to have been normalized so as to sum to one. These are divided by the standard population rate $\sum w_j \lambda_j^*$ in order to find the CMFs for each level. In the examples below, the standard weights depend only on age (Table 3.2).

Example 3.1

Table 3.3 illustrates the application of several elementary methods to the grouped data from the Montana study. Crude and directly standardized death rates are shown in the first few rows of each part of the table, the two parts corresponding to the pre- and post-1925 subcohorts created to illustrate the effect of date of hire. Stratum-specific death rates for each exposure group, calculated from the deaths and person-years in Appendix V, were multiplied by the standard weights (Table 3.2) and summed to give the directly standardized rate. The standard population rate used for comparison is simply the weighted average of the stratum-specific standard rates, the same weights being used for each calendar period. Both sets of standardized rates were divided by the total of the weights for those age × calendar periods that had some person-years of observation for the particular exposure category. In the second part of Table 3.3, note the limitation in the use of the standardized rate as a comparative measure caused by a lack of data for certain age × calendar periods for persons with the longest exposure. There is a substantial jump in the standardized rates as one progresses from the first to the second exposure category, but a less obvious trend thereafter.

In actual epidemiological practice, examination of dose-response trends in terms of directly standardized rates or CMFs seems largely and properly to be limited to studies in which there are substantial numbers of deaths in each exposure category. This ensures that the standardized rates are reasonably stable, so that evidence for a trend should be clear from a simple examination of the data. Although questions of statistical significance are generally not at issue, it is nonetheless prudent to report the standard

Table 3.5 Age-standardized death rates (all causes combined) and mortality ratios among male ex-smokers of 1–19 cigarettes per day, ages 50–74 years[a]

Smoking category	Number of men	Number of deaths	Standardized death rate (per 100 000 person-years)	CMF
Current smokers	118 373	9 117	2 359	172
Ex-smokers by years since last smoked				
Under 1	814	64	2 212	161
1–4	1 986	144	1 985	144
5–9	1 909	128	1 840	134
10+	4 578	255	1 397	102
Nonsmokers	62 332	3 512	1 374	100

[a] From Hammond (1966)

error of each summary rate (equation (2.7)) as a means of judging its stability. In case of uncertainty about the statistical significance of the observed results, the reciprocals of the corresponding variances could be employed as empirical weights in a formal regression analysis of the directly standardized rates on quantitative exposure variables. Such a regression analysis could also be helpful if the summary data were the only data available, for example, if they were obtained from published sources. However, statisticians have pointed out the need for caution in regression analyses of standard rates or other indices that have (age-specific) population denominators in both dependent and independent variables.

Example 3.2

The American Cancer Society study of one million men and women (Hammond, 1966) furnishes an example in which numbers of deaths are sufficiently large that direct standardization is appropriate. The effect of smoking on mortality was reported in terms of the ratios of the standardized death rates for various categories of smokers relative to the standardized rate for nonsmokers. Table 3.5, which concerns smokers of 1–19 cigarettes per day, indicates that cessation of smoking for increasing lengths of time results in a decline in the all-causes death rate compared to that for continuing smokers. Ten years after cessation of exposure, the death rate among ex-smokers is down nearly to the level among lifelong nonsmokers.

3.4 Comparison of standardized mortality ratios

If the data are not so extensive and questions of sampling variability are of greater concern, it is generally appropriate to use the SMR in place of the CMF as a measure of how the death rates in each exposure category compare with those of the standard population. Evidence for a dose-response trend may then be sought in terms of an increase or decrease in the SMRs with increasing exposure. Referring to Table 3.4, let us denote by $O_k = \sum_j d_{jk}$ the observed number of deaths in the kth exposure group. Keeping to the convention that quantities calculated from external standard rates are starred (*), the expected numbers of deaths may be written

$$E_k^* = \sum_j n_{jk}\lambda_j^* \tag{3.2}$$

and the standardized mortality ratios

$$\mathrm{SMR}_k = O_k / E_k^* . \tag{3.3}$$

The grand totals of observed and expected deaths are denoted by $O_+ = \sum O_k$ and $E_+^* = \sum E_k^*$, respectively.

The overall SMR for the entire cohort is given by O_+/E_+^*, which was discussed in detail in Chapter 2. Here we are interested in comparisons among the different subcohorts, that is, among the different SMR_ks. When examining the SMR_ks for a trend with increasing exposure, it should be kept in mind that they are relative measures of effect calculated with reference to an external set of rates and that they may not be strictly comparable to one another. For reasons discussed at length in §2.3, ratios of the SMR_ks for different exposure categories may fail to summarize adequately the ratios of the stratum-specific rates. This occurs in precisely those circumstances when the SMR_ks themselves are not good summary measures, namely when the ratios of cohort to standard death rates vary widely from one stratum to another. For example, it might happen that heavier exposures had the effect of adding progressively greater amounts to the age-specific (background) rates that would be expected in the absence of exposure. However, if much higher background rates were expected with the heavier exposures, for instance, because persons with such exposures tended to be older, such an additive dose-response relationship could well be missed by a comparison of SMR_ks.

Example 3.3

Table 3.6 presents fictitious data that illustrate the phenomenon just described. The effect of increasing exposure is to increase the two age-specific death rates by 2 per 100 person-years (low exposure) or 4 per 100

Table 3.6 Fictitious data to illustrate a potential defect in the SMR

	Age range (years)			SMR (%)	CMF (%)
	35–44	45–54	Total		
Unexposed					
Rate (per 100)	3	10	3.7	176	100
Population	900	100	1000		
No. of observed deaths	27	10	37		
No. of expected deaths	9	12	21		
Lightly exposed					
Rate (per 100)	5	12	8.5	131	131
Population	500	500	1000		
No. of observed deaths	25	60	85		
No. of expected deaths	5	60	65		
Heavily exposed					
Rate (per 100)	7	14	13.3	121	161
Population	100	900	1000		
No. of observed deaths	7	126	133		
No. of expected deaths	1	108	109		
Standard population					
Rate	1	12	6.5		
Weight	0.5	0.5	1.0		

person-years (high exposure). These increases are reflected in increases in the CMFs, which are averages of the age-specific rates. However, due to the very skewed age distribution and the fact that the cohort to standard rate ratios vary markedly with age, the apparent trend as measured by the three SMR_ks is reversed. Compare Table 2.13.

Fortunately, the statistical confounding is not so serious in typical applications, and a dose-response analysis carried out in terms of SMR_ks often yields results that are quite similar to those obtained by other methods. The formal assumption required for an SMR analysis to be completely appropriate is that the stratum-specific death rates for each exposure class be proportional to the external standard rates, this being precisely the condition needed to assure comparability of the SMR_ks. This assumption may be investigated in practice by fitting an explicit model and comparing the observed and fitted number of deaths in each stratum-exposure cell, using the techniques described in the next chapter. Thus, the data themselves should give indications of situations in which inferences based on the SMR are liable to be seriously in error. When there appears to be heterogeneity of (multiplicative) dose-response effects between different age strata, it is better to use a different model to describe this heterogeneity rather than to summarize a number of disparate effects in a single SMR.

When the proportionality assumption holds, we may regard the total number of deaths, O_k, observed at the kth exposure level as having an approximate Poisson distribution with mean $\theta_k E_k^*$, where E_k^* represents the expected number of deaths and θ_k the unknown SMR for this level of exposure in relation to the standard rates (see §4.3). The ratios of SMR_ks, which we denote $\psi_k = \theta_k/\theta_1$, thus represent relative risks for each exposure level using the first level as baseline ($\psi_1 = 1$). These have precisely the same interpretation as do the relative risk parameters ψ_k estimated in case-control studies (§4.5 of Volume 1). They represent the ratios of age-specific rates for different exposure categories, assuming these to be constant over age-calendar year strata.

In this section we consider methods for estimating the individual relative risks, for determining their standard errors, for testing the statistical significance of each one individually, and for testing the global null hypothesis that the ψ_k equal unity (i.e., the SMR_ks are equal) for $k = 1, \ldots, K$ against alternatives of heterogeneity and trend. The tests involve a comparison of the observed numbers O_k with fitted values $\tilde{E}_k^*$ calculated under the hypothesis that each θ_k is equal to some common value θ. These latter are easily obtained by distributing the total deaths O_+ among the K exposure levels in proportion to the expected numbers:

$$\tilde{E}_k^* = O_+(E_k^*/E_+^*). \tag{3.4}$$

We refer to the $\tilde{E}_k^*$ as 'adjusted expected values' to reflect the fact that they are equal to the E_k^* scaled by the overall SMR, O_+/E_+^*, so as to ensure that $\sum_k \tilde{E}_k^* = \sum_k O_k$.

(*a*) *Two dose levels*: *exposed* versus *unexposed*

The simplest comparison is between two levels of exposure, say exposed ($k = 2$) *versus* unexposed ($k = 1$). Thus, we regard O_1 as a Poisson variable with mean $\theta_1 E_1^*$ and O_2 as Poisson with mean $\theta_2 E_2^*$. If we set $\theta = \theta_1$, $\psi = \psi_2 = \theta_2/\theta_1$ and $\theta\psi = \theta_2$, suppressing the subscripts for clarity, the parameter of interest is the relative risk ψ, θ

playing the role of a nuisance parameter that interferes with our inferences concerning ψ. According to standard principles of statistical inference (Cox & Hinkley, 1974), it is appropriate in such circumstances to consider a distribution for the observed data which depends only on the parameter of interest. This is quite easy here since the distribution of two Poisson variates conditional on their sum is binomial (Lehman, 1959). More precisely,

$$pr(O_2|O_+;\theta,\psi)=\binom{O_+}{O_2}\pi^{O_2}(1-\pi)^{O_1}, \tag{3.5}$$

where

$$\pi=\frac{\psi E_2^*}{E_1^*+\psi E_2^*}$$

or, equivalently,

$$\psi=\frac{\pi E_1^*}{(1-\pi)E_2^*}. \tag{3.6}$$

Statistical inferences about the relative risk ψ, whether exact or approximate, may therefore be carried out by making inferences about the binomial parameter π in (3.5) and then transforming *via* (3.6). They are formally identical to those used in the analysis of matched case-control pairs with dichotomous exposures (§5.2 of Volume 1). The relevant equations need merely be rewritten for use with cohort data.

Under the null hypothesis $\psi_0=1$ we have $\pi_0=E_2^*/E_+^*$, and an exact test is obtained from the tail probability of the corresponding binomial distribution. For example, if $O_2>E_2^*$, the one-sided significance level or p value is given by

$$p=\sum_{x=O_2}^{O_+}\binom{O_+}{O_2}\pi_0^x(1-\pi_0)^{O_+-x}.$$

In practice, it will usually suffice to use the approximate chi-square statistic based on the observed deviation of O_2 from its expectation. This may be written

$$\chi^2=\frac{\{|O_2-E(O_2)|-\frac{1}{2}\}^2}{\mathrm{Var}(O_2)}=\frac{\{|O_1-\tilde{E}_1^*|-\frac{1}{2}\}^2}{\tilde{E}_1^*}+\frac{\{|O_2-\tilde{E}_2^*|-\frac{1}{2}\}^2}{\tilde{E}_2^*}, \tag{3.7}$$

where we have used the fact that $\mathrm{Var}\,(O_2)=O_+\pi_0(1-\pi_0)=\tilde{E}_1^*\tilde{E}_2^*/(\tilde{E}_1^*+\tilde{E}_2^*)$ and $O_1-\tilde{E}_1^*=-(O_2-\tilde{E}_2^*)$. The numerator(s) in (3.7) are reduced in absolute value by 1/2 before squaring for a continuity correction.

(*b*) *Point and interval estimation of the relative risk*

The maximum likelihood estimate of π is $\hat{\pi}=O_2/O_+$, from which it follows that the maximum likelihood estimate of ψ is

$$\hat{\psi}=\frac{\hat{\pi}\tilde{E}_1^*}{(1-\hat{\pi})\tilde{E}_2^*}=\frac{O_2E_1^*}{O_1E_2^*}, \tag{3.8}$$

the ratio of the two SMR_ks. Exact $100(1-\alpha)\%$ confidence limits for π may be found from the charts of Pearson and Hartley (1966) or computed using the equations

$$\pi_L = \frac{O_2}{O_2 + (O_1 + 1)F_{\alpha/2}(2O_1 + 2, 2O_2)}$$

and (3.9)

$$\pi_U = \frac{(O_2 + 1)F_{\alpha/2}(2O_2 + 2, 2O_1)}{O_1 + (O_2 + 1)F_{\alpha/2}(2O_2 + 2, 2O_1)}$$

where $F_{\alpha/2}(v_1, v_2)$ denotes the upper $100_{\alpha/2}$ percentile of the F distribution with v_1 and v_2 degrees of freedom. The limits (3.9) are inserted into (3.6) to obtain confidence limits ψ_L and ψ_U for the relative risk. Alternatively, approximate limits based on the normal approximation to the binomial probabilities (Cornfield, 1956) are given as the solutions to the equations

$$E_1^*(O_2 - 1/2) - \psi_L E_2^*(O_1 + 1/2) = +Z_{\alpha/2}\{\psi_L E_1^* E_2^*(O_1 + O_2)\}^{1/2}$$

and (3.10)

$$E_1^*(O_2 + 1/2) - \psi_U E_2^*(O_1 - 1/2) = -Z_{\alpha/2}\{\psi_U E_1^* E_2^*(O_1 + O_2)\}^{1/2}.$$

These are quadratic equations in the unknown variable $\xi = \sqrt{\psi}$.

Example 3.4

Suppose $O_1 = 5$ and $O_2 = 14$ bladder cancer deaths are observed among unexposed and exposed members of an industrial cohort, respectively, whereas $E_1 = 7.3$ and $E_2 = 5.5$ were expected from vital statistics available for the region in which the plant was located. The overall SMR is $O_+/E_+ = 19/12.8 = 1.484$, and adjusted expected values are $\tilde{E}_1^* = 7.3 \times 1.484 = 10.84$ and $\tilde{E}_2^* = 5.5 \times 1.484 = 8.16$. Individual SMRs are $5/7.3 = 0.685$ and $14/5.5 = 2.545$ for unexposed and exposed so that $\hat{\psi} = 2.545/0.685 = 3.72$ is the point estimate of relative risk. The test (3.7) for the hypothesis $\psi = 1$ gives

$$\chi^2 = \frac{(5 - 10.84 + 0.5)^2}{8.16} + \frac{(14 - 8.16 - 0.5)^2}{10.84} = 6.13$$

with continuity correction ($p = 0.01$). Using equation (3.9) and the fact that $F_{0.025}(12, 28) = 2.45$ and $F_{0.025}(30, 10) = 3.31$, exact 95% confidence limits on the associated binominal probability are

$$\pi_L = \frac{14}{14 + 6(2.45)} = 0.488$$

and

$$\pi_U = \frac{15(3.31)}{5 + 15(3.31)} = 0.908,$$

from which we determine $\psi_L = 1.26$ and $\psi_U = 13.2$ as limits on the relative risk. The approximate limits are found as solutions to

$$98.55 - \psi_L 30.25 = 1.96\{\psi_L 762.85\}^{1/2}$$

and

$$105.85 - \psi_U 24.75 = -1.96\{\psi_U 762.85\}^{1/2},$$

these being $\psi_L = 1.25$ and $\psi_U = 11.8$, respectively.

(c) *Testing for heterogeneity and trend in the SMRs*

The same methods may be used to estimate the relative risks ψ_k for each level of exposure $k = 2, \ldots, K$ and to test the significance of each one individually. One merely substitutes O_k and $\tilde{E}_k^*$ for O_2 and $\tilde{E}_2^*$ in equations (3.7) through (3.10). However, since interpretation of a large number of separate comparisons is difficult, we also need a test of the hypothesis that all K ψ_k are simultaneously equal to unity. This is easy to derive using the framework already introduced (Kilpatrick, 1962, 1963). Conditional on the total observed deaths, O_+, the joint distribution of $\boldsymbol{O} = (O_1, \ldots, O_K)$ under the null hypothesis is multinomial with cell occupancy probabilities $(\pi_1, \ldots, \pi_K)$ where $\pi_k = E_k^*/E_+^*$. A test of the global null hypothesis is thus achieved by comparing the O_k to the fitted values $\tilde{E}_k^*$ using the standard criterion

$$\chi^2_{K-1} = \sum_{k=1}^{K} \frac{(O_k - \tilde{E}_k^*)^2}{\tilde{E}_k^*}, \tag{3.11}$$

which should be referred to tables of the chi-square distribution with $K - 1$ degrees of freedom.

One disadvantage of (3.11) is its relative lack of power against the specific alternative hypothesis of a trend in the SMR_ks with increasing exposure. Even if none of the pairwise comparisons of baseline and exposure groups nor the multi-degree of freedom statistic (3.11) yields a significant result, substantial evidence for a dose-response trend may nevertheless be generated if the estimated relative risks are in the hypothesized order. The Poisson trend statistic (Armitage, 1955; Tarone, 1982) was designed especially to detect such monotonic dose-response relationships. If x_k denotes a quantitative dose level associated with the kth exposure category, this single degree of freedom test is given by

$$\chi_1^2 = \frac{\{\sum_{k=1}^{K} x_k(O_k - \tilde{E}_k^*)\}^2}{\sum_{k=1}^{K} x_k^2 \tilde{E}_k^* - (\sum_{k=1}^{K} x_k \tilde{E}_k^*)^2/O_+}. \tag{3.12}$$

In situations in which the categories are merely ordered, and there is no specific quantitative exposure, it suffices to set $x_k = k$. Formal justification for both (3.11) and (3.12) stems from the fact that they are efficient score tests under various sets of assumptions, including the log linear models for Poisson variables discussed in the next chapter (Tarone & Gart, 1980).

Example 3.5

Returning to the Montana data, the standard rates shown in Table 3.2 were used in conjunction with the data in Appendix V to produce expected numbers of deaths and SMR_ks for each exposure category using equations (3.2) and (3.3). With the exception of the highest dose category for the post-1924 cohort, the SMR_ks are in reasonable agreement with the corresponding CMF_ks (Table 3.3). However, the CMF for this category is not comparable to the others since there are no data for the earliest calendar period for two age groups. To alleviate this difficulty, we could, of course, restrict all CMFs to those age × calendar period strata for which full data are available. Relative risks obtained by dividing each SMR_k by the SMR for 0–0.9 years exposed indicate that workers hired before 1925 who had 15 or more years of moderate to heavy arsenic exposure have mortality rates from respiratory cancer that are approximately three times higher than the rates among workers who remained in areas of the plant where only light exposures occurred.

The penultimate rows in both parts of Table 3.3 show the adjusted expected values $\tilde{E}_k^*$ for the four

exposure categories. These were obtained in accordance with equation (3.4), multiplying the expected numbers shown in the sixth row of each part of the table by the overall SMR. In the 0 – 0.9 years exposure group for pre-1925 employees, for example, we have $21.47 \times (115/31.62) = 78.10$ cases expected after adjustment. The global test (3.11) yields

$$\chi_3^2 = \frac{(51-78.10)^2}{78.10} + \frac{(17-10.71)^2}{10.71} + \frac{(13-10.02)^2}{10.02} + \frac{(34-16.17)^2}{16.17} = 33.7$$

as shown at the bottom of the first part of Table 3.3. Likewise, the trend statistic (3.12) is

$$\chi_1^2 = \frac{\{1\times(51-78.10)+2\times(17-10.71)+3\times(13-10.02)+4\times(34-16.17)\}^2}{(1\times 78.10+4\times 10.71+9\times 10.02+16\times 16.17)-(78.10+2\times 10.71+3\times 10.02+4\times 16.17)^2/115} = 30.5$$

Note the use of the coded levels $x_k = k$ in this example.

(*d*) *Trend test for exposure effect* versus *trend test for dose-response*

The object of a dose-response analysis is to demonstrate a continuously increasing response to increasing dose or, in the present context, a continuously increasing (relative) risk with increasing exposure. While the trend statistic (3.12) is designed to detect such alternatives to the null hypothesis (no effect of exposure), it may sometimes give a significant result even if the relative risks are not continuously increasing. This could happen, for example, if the risk were increased for any amount of exposure relative to no exposure but the risks among the different exposure levels remained constant. The causal inference linking exposure and disease is less secure in such cases, because of the greater possibility that a dose-response function that jumps up initially and then remains flat could be produced by bias or confounding. For example, the weak relationship between coffee drinking and bladder cancer observed in several case-control studies was interpreted as noncausal on just such a basis (§3.2 of Volume 1). One may wish to restrict the trend statistic to a comparison of positive dose or duration levels and exclude the baseline nonexposed category when testing specifically for a dose-response effect.

Example 3.6

Returning to Table 3.3, we noted significant 'trends' in relative risk with increasing duration of heavy/medium arsenic exposure for both the pre-1925 and post-1925 sub-cohorts ($\chi_1^2 = 30.5$ and 8.74, respectively). However, the relative risk estimates in fact showed little variation among the three highest categories of exposure. Restricting the trend analyses to the categories 1.0–4.9 years duration, 5.0–14.9 years, and 15+ years, this being accomplished by adjusting the expected values to agree with the total observed for the three categories, and applying the usual statistic (3.12), we find $\chi_1^2 = 1.67$ ($p = 0.19$) for the pre-1925 cohort and $\chi_1^2 = 0.60$ ($p = 0.44$) for the post-1925 cohort. This confirms what is already apparent from an examination of the relative risks, namely, that there is no evidence for an increasing dose-response trend with exposure duration above one year.

If the object of the analysis is primarily to test for a possible carcinogenic effect, however, the baseline or lowest dose level should definitely be included in calculation of the trend. An issue that then arises is whether the intercept of the regression line of SMR on dose, the slope of which is implicitly being tested in a trend analysis, necessarily passes through unity or instead through some other value that represents the true position of the cohort *vis-á-vis* the standard population. If the true SMR at zero dose were somehow known *a priori* to be equal to one, although this is unlikely in

practice, this would permit a more powerful analysis and yield a more significant result on average (Gilbert, 1983).

The trend statistic (3.12) implicitly assumes that the intercept is being estimated from the data. One argument in favour of estimating the intercept is that the cohort may have higher death rates than expected even in the low dose range, due to the effects of other risk factors. Or the initial SMR may be less than 1 due to a 'healthy worker' selection effect. A trend analysis that assumed it was equal to 1 would yield a trend test statistic that was too large in the first case and too small in the second. With the Montana study, for example, the SMR_ks for respiratory cancer in the lowest dose groups are 237.5% and 134.9% for the pre- and post-1925 cohorts, respectively (Table 3.3). It is unclear whether the excess is due to generally higher levels of smoking in the study population or to the effects of arsenic exposures that even 'low dose' persons may receive. Thomas, D.C. and McNeill (1982) note that other reasons for the regression line not to pass through unity at zero dose, besides the possible noncomparability of the standard population, are that the assumed dose-response function is wrong or that random errors in dose measurement have led to a slope estimate that is too shallow. In the face of such uncertainty, it does not seem prudent to make a strong assumption about the intercept.

(*e*) *Selection of the dose metameter*

In order to carry out the test for trend we must assign quantitative values to each exposure category. Underlying the test is the implicit assumption that some transformation of the disease rate is a linear function of a dose variable x. It is the slope of this relation that is being tested (Tarone & Gart, 1980). Thought should be given to the most appropriate values, since the choice sometimes can have a substantial influence on the significance of the result. Often one will want to choose the dose scale so that there is an approximately linear relationship between disease rates and exposures, at least at low doses. Multistage models of carcinogenesis (Chapter 6) suggest a low-order polynomial relationship of the form $\lambda(d) = \beta_0 + \beta_1 d + \beta_2 d^2 + \ldots$, where all coefficients are positive. These imply that there is an approximately linear relationship between relative risk and exposure at low (measured) doses. Other assignments of x values to exposure categories may be tried also, although problems of interpretation will arise if a large number of separate tests are carried out on the same data.

Example 3.7

Table 3.7 shows numbers of deaths from haematological malignancies among workers at the Portsmouth (US) Naval Shipyard, according to the cumulative radiation exposure received by the time of death (Najarian, 1983). Also shown are person-years denominators by dose category. In order to test for a trend in the Poisson rates with increasing dose, we use Armitage's (1955) statistic, which has the same form as (3.12) except that the expected deaths $\tilde{E}_k$ are obtained by allocating the total deaths in proportion to the person-years in each category. Both observed and expected deaths are shown in Table 3.7.

Note that the intervals used for grouping the data into exposure categories are approximately logarithmic. After the initial control category, each group of radiation doses is approximately ten times larger than the preceding one. Thus, the usual assignment of coded values $x_k = k$ to the $K = 7$ exposure categories effectively means that a log dose metameter is being used. An alternative would be to use a linear dose metameter, assigning to each exposure category the average of the doses included within it. Someone who believed that the true dose-response curve was discontinuous and that there was a threshold at 1 rad might

Table 3.7 Observed and expected deaths from haematological malignancies among Portsmouth (USA) Naval Shipyard workers, by cumulative radiation dose[a]

Lifetime dose (rems)	No. of observed deaths	No. of person-years	No. of expected deaths[b]	Dose metameters		
				Linear	Log	Threshold
0.000	2	24 232	3.36	0.0	1	0
0.001–0.030	2	14 810	2.06	0.015	2	0
0.030–0.100	0	15 960	2.22	0.065	3	0
0.100–0.500	4	24 138	3.35	0.30	4	0
0.500–1.00	1	10 418	1.45	0.75	5	0
1.00–5.00	7	21 769	3.02	3.0	6	1
5.00+	1	11 128	1.54	6.0	7	1
Total	17	122 455	17.00			
Test for trend (χ^2_1):				1.19	2.25	3.53
p value (one-sided):				0.14	0.07	0.03

[a] From Najarian (1983)
[b] Assuming constant rate in all dose groups

assign $x_k = 0$ for $k = 1, 2, 3, 4, 5$, and $x_k = 1$ for $k = 6$ and 7. We would have to be suspicious of this choice for the threshold, however, since setting it at 1 rad for these particular data obviously maximizes the difference in relative risks one will observe between the 'exposed' ($x_k = 1$) and 'unexposed' ($x_k = 0$).

The three dose metameters lead to rather different trend statistics in this example. The logarithmic scale yields $\chi^2 = 2.25$ ($p = 0.07$; one-sided), whereas the value on the arithmetic scale is somewhat lower at $\chi^2 = 1.19$ ($p = 0.14$). The most significant result is obtained from the threshold model comparing doses over and under 1 rad ($\chi^2 = 3.53$, $p = 0.03$). Setting the threshold at 0.5 rads reduces χ^2 from 3.53 to 2.32, indicating the sensitivity to a basically arbitrary threshold. Since the results obtained with the continuous scales do not attain statistical significance, one would conclude little more than that the situation perhaps warranted further investigation. No excess of deaths due to cancer nor specifically to cancer of the blood or blood-forming tissues was found in the analysis of these data performed by Rinsky *et al.* (1981). It seems likely that the positive results reported from the earlier proportional mortality study (Najarian & Colton, 1978) were biased by the incomplete ascertainment of deaths that had occurred among workers at the facility (Committee on the Biological Effects of Ionizing Radiation, 1980). See also the discussion in §1.6.

(*f*) *Alternative tests for trend*

The statistic (3.12) that we have suggested for a trend test relies heavily on the assumed Poisson variability of the observed numbers of deaths. In some situations in which there are a large number of different comparison groups, it may be more prudent to carry out a standard regression analysis of the SMR_ks or their logarithms on the quantitative dose levels. Especially when the SMR_ks are calculated for different intervals of calendar time, the observed variation in numbers of deaths between adjacent time intervals may be greater than would be expected from Poisson sampling variation. It is then more appropriate to evaluate the linear time trend against the observed background of year-to-year variation rather than against the smaller theoretical variance. The issue is complicated by the fact that the estimate of residual variation from the regression analysis may be heavily dependent on the particular regression model chosen, or may be unstable because the number of different dose

categories does not provide a sufficient number of 'degrees of freedom' for error estimation. Furthermore, there is some controversy regarding the extent to which one should account for the underlying Poisson variability by giving greater weight in the regression analysis to SMR_ks based on large numbers of deaths.

The ideal solution is probably intermediate between an unweighted analysis, in which most of the observed variability is attributed to extraneous factors rather than sampling, and an analysis based entirely on Poisson sampling theory (Pocock *et al.*, 1981; Breslow, 1984a). A practical alternative is to carry out both Poisson and unweighted regressions and compare results.

Example 3.8

Table 3.8 presents data from the study of Rocky Mountain uranium miners quoted by Thomas, D.C. and McNeill (1982). There is a reasonably linear relationship between the logarithm of the SMR and the logarithm of the average cumulative radiation exposure, measured in working level months (WLM) (Fig. 3.2). As noted in equation (2.9), the (Poisson) variance of the log SMR is estimated approximately by the reciprocal of the number of deaths, which suggests we use the number of deaths to weight the individual observations. A weighted linear regression analysis of log SMR on log WLM yields a residual (weighted) sum of squares of 6.68 on six degrees of freedom. We conclude that the extra-Poisson variability in this case is minor or nonexistent, since, otherwise, the residual mean square would be substantially larger than one. The corresponding F statistic for the significance of the linear trend is 57.4 on one and six degrees of freedom. An unweighted analysis yields $F_{1,6} = 59.8$. Both results confirm the highly significant $\chi^2_1 = 95$ that is found from the usual trend test (3.12).

If one considers instead a linear regression of the SMR_k on dose x_k, weighting each observation by $(E_k^*)^2/O_k$, where E_k^* is the expected and O_k the observed number of deaths, the residual mean square is $11.09/6 = 1.85$. In view of the preceding results, the excess above unity is probably due more to the lack of fit of the linear model than to non-Poisson variation. The F statistics are 41.9 for the weighted analysis and 261.4 for the unweighted. The discrepancy between the weighted and unweighted test statistics on the arithmetic scale results from the data point for the highest dose category being far removed from the others and having a much greater influence on the unweighted analysis than the weighted one. This instability reminds us of the dangers of the uncritical use of least-squares regression techniques, especially with small samples, and suggests that they are best reserved for situations in which there is a large number of dose categories. Alternatively, modern techniques of robust regression (Huber, 1983) may be used.

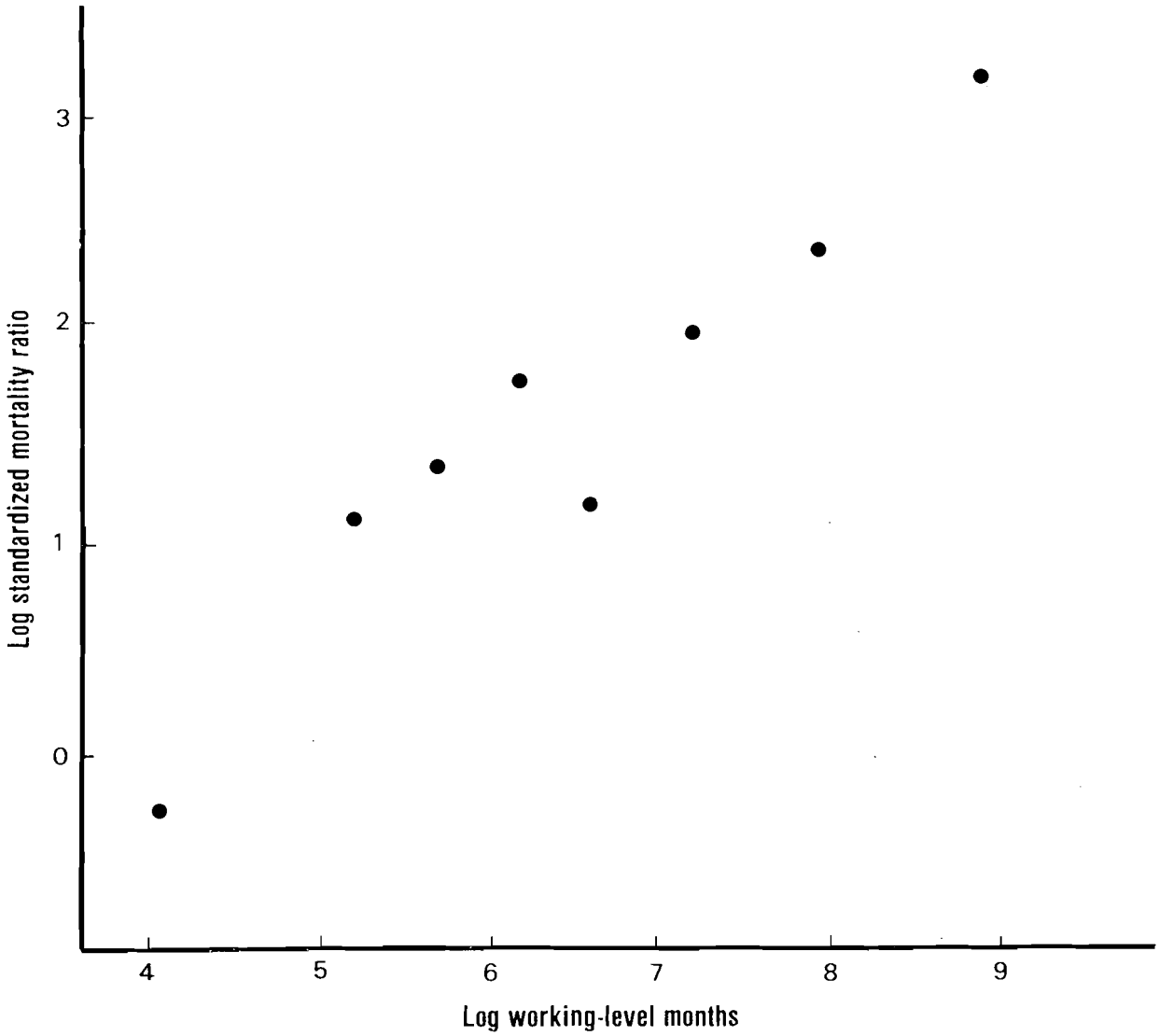
Table 3.8 Lung cancer risk in US uranium miners[a]

Cumulative WLM[b]		Person-years	Lung cancers		SMR (%)
Range	Midpoint		Observed	Expected	
0–119	60	5 183	3	3.96	76
120–239	180	3 308	7	2.24	312
240–359	300	2 891	9	2.24	402
360–599	480	4 171	19	3.33	571
600–839	720	3 294	9	2.62	344
840–1 799	1 320	6 591	40	5.38	743
1 800–3 719	2 760	5 690	49	4.56	1 075
>3 719	7 000 (est)	1 068	23	0.91	2 727
All	1 180 (mean)	32 196	159	25.24	

[a] From Committee on the Biological Effects of Ionizing Radiation (1980) as quoted by Thomas, D.C. and McNeill (1982)

[b] WLM, working-level-month measure of cumulative exposure

Fig. 3.2 Log–log plots of SMRs for US uranium miners from Table 3.8

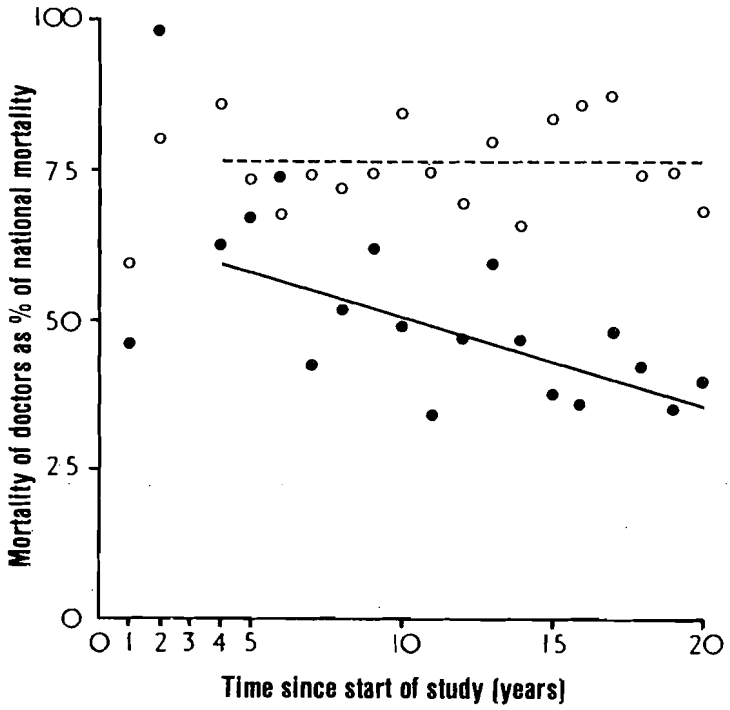


(g) Some examples from the literature

Doll and Peto (1976) reported results of the 20-year follow-up of British doctors to study cigarette smoking and mortality. Most of their analyses compared cause-specific mortality rates among exposure categories determined by smoking history, using methods of internal standardization that are described below. However, the authors also wanted to see whether the fact that doctors gave up smoking more rapidly than members of the general population was reflected in an improvement in their relative survival. Thus, mortality rates for all of England and Wales were used as a standard for computation of SMR_ks for each calendar year for two causes of death (Fig. 3.3). The evident decline in the relative rates of lung cancer was confirmed by a least-squares linear regression analysis of the 20 SMR_ks on calendar year.

Another example illustrates more specifically the use of the Poisson trend statistic. Table 3.9 presents leukaemia mortality rates during various intervals following first

Fig. 3.3 Trend in number of deaths certified in British male doctors as percentage of number expected from experience of all men in England and Wales of the same ages. Results are given from the second to the twentieth years of study for lung cancer (●) (459 deaths observed *versus* 931.9 expected) and all other cancers (○) (1238 deaths observed *versus* 1630.7 expected). Regression lines on time were calculated from data for the fourth to the twentieth years of study (regression coefficients: −1.4 for lung cancer and 0.0 for all other cancers). From Doll and Peto (1976)



treatment for a cohort of ankylosing spondylitis patients (Smith & Doll, 1982). The expected numbers shown were also obtained from mortality rates for England and Wales specific for sex, age and calendar year. In this example, the statistic (3.12) gives a value of $\chi_1^2 = 10.40$ and provides clear evidence for a decline in the observed: expected ratios with increasing time since exposure.

Finally, Table 3.10 presents data from a cohort study of US and Canadian insulation

Table 3.9 Observed and expected leukemia deaths among ankylosing spondylitis patients, by time since initial treatment[a]

	Time since treatment (years)							Total
	0–2	3–5	6–8	9–11	12–14	15–17	18+	
Observed	6	10	6	3	1	4	1	31
Expected	1.00	0.89	0.87	0.90	0.96	0.90	0.95	6.47
SMR	6.00	11.24	6.90	3.33	1.04	4.44	1.05	4.79

[a] From Smith and Doll (1982)

Table 3.10 Lung cancer deaths and person-years among asbestos and insulation workers according to duration of time since initial exposure[a]

Duration (years)	Number of men	Person-years	No. of observed deaths	No. of expected deaths	SMR (%)
0–9	8 190	26 393	0	0.7	—
10–14	9 063	29 003	7	2.7	255
15–19	9 948	34 066	29	8.5	340
20–24	8 887	31 268	59	17.0	348
25–29	6 596	20 657	105	21.0	500
30–34	3 547	11 598	112	18.4	608
35–39	2 020	5 403	65	11.5	568
40–44	1 108	3 160	40	8.1	493
45+	1 448	5 305	69	17.8	389

[a] From Selikoff *et al.* (1980)

workers (Selikoff *et al.*, 1980). Ratios of observed to expected lung cancer deaths reached a peak between 30–35 years from the initial exposure to asbestos. This does not mean, of course, that the absolute rates of lung cancer decline after 35 years, although this is a common misconception. The death rates continue to increase as the exposed workers grow older, but at a slightly lower rate in comparison to the general population than was true during earlier years. Because the SMRs first rise and then fall, one could well expect the trend statistic not to yield a significant result in this example. Various possible explanations have been suggested for the decline. One is that the combined exposure to asbestos and cigarettes was so lethal that heavy smokers were eliminated from the study cohort at an even faster rate than they were eliminated from the general population. Another possibility is that the termination of exposure following retirement, which would start to occur 35 years or so after initial employment, led to an attenuation of subsequent relative risk but at a much slower pace than that noted for ex-smokers (Table 3.5). Thirdly, it should be noted that there is a strong confounding effect in this cohort between period of initial exposure, when different types of asbestos fibres may have been used or the exposure intensity different, and the time since first exposure. Finally, the SMR_ks reflect any difference in smoking patterns between asbestos workers and the general population, and these also may have been changing over time.

3.5 Comparison of internally standardized mortality ratios

The methods of analysis discussed so far rely on standard rates that are external to the study cohort in order to make comparisons between exposure groups. Questions about the appropriateness of the particular standard selected and the comparability of the resulting SMR_ks suggest that a more satisfactory approach would be to use the observed data, without consideration of any outside rates, when making internal comparisons.

From a theroretical viewpoint, the method of internal standardization is probably

best regarded as a rough and ready approximation to the more complicated but more appropriate methods of grouped data analysis that are presented in the next section. If there are only two exposure categories, it tends to yield mildly conservative tests and estimates in typical practice (Bernstein *et al.*, 1981; see also Fig. 4.3). The conservatism could be substantial if age and calendar time or other stratification variables strongly confound the exposure-disease relationship. Nevertheless, the method of internal standardization enjoys a considerable following due to its relative simplicity and strong intuitive appeal.

If there are more than two exposure categories, internal standardization does not eliminate the problem that was discussed at length in §2.3 concerning the comparability of SMRs. Although the external standard is replaced by an internal standard consisting of the combination of all exposure groups, in particular examples this pooled group may be dominated by one or two large exposure groups. When comparing the ratios of SMR_ks for two other exposure groups, therefore, it is possible for the same type of bias to occur.

The calculations required for internal standardization are surprisingly easy. Referring to the data layout in Table 3.4, the stratum-specific death rates calculated without regard to exposure category are $\lambda_j = D_j/N_j$. It follows that the expected number of deaths in the kth exposure class, assuming that exposure had no effect on the rates, is

$$E_k = \sum_{j=1}^{J} n_{jk}\lambda_j = \sum_{j=1}^{J} n_{jk}D_j/N_j. \tag{3.13}$$

These internally derived fitted values share with the adjusted expected numbers (3.4) the property that their sum is equal to the total number of observed deaths. They are used in place of the $\tilde{E}_k^*$ in equations (3.7), (3.8), (3.11) and (3.12) in order to make approximate estimates of the relative risks for each exposure category and approximate tests of their heterogeneity and trend. As already noted, these tests and estimates tend to be somewhat conservative, more so if there is a high degree of association between the stratum variables and the exposures. However, this feature is not well illustrated by the data on the Montana workers, since, as often happens in practice, the degree of confounding is rather slight.

Example 3.9

By pooling the respiratory cancer deaths and person-years shown in Appendix V over period of hire and duration of exposure, one obtains the pooled death rates shown in Table 2.8 by ten-year intervals of age and calendar period. Table 3.11 presents the expected numbers of deaths calculated for each exposure category by multiplying the pooled rates by the appropriate number of person-years and summing in accordance with equation (3.13). Separate analyses were carried out according to period of employment. Note the similarity between the internally fitted values and the adjusted expected values shown in Table 3.3. The latter are slightly more extreme and therefore indicate a slightly steeper dose-response relationship. For example, the estimated relative risk for the highest exposure category among those employed prior to 1925 is 3.22 for external standardization *versus* 3.09 for internal standardization.

Inserting the observed and expected values from Table 3.11 in equations (3.11) and (3.12), following exactly the same method of calculation as in Example 3.5, the values of the tests for heterogeneity and trend are $\chi_3^2 = 31.7$ and $\chi_1^2 = 28.3$, respectively, for the pre-1925 subgroup. These are less than the values found with external standardization, but they are still highly significant ($p < 0.0001$). A similar result holds for the post-1925 subgroup.

Table 3.11 Dose-response analysis of respiratory cancer deaths among Montana smelter workers, based on internal standardization

	Cumulative years of moderate/heavy arsenic exposure				
	0–0.9	1.0–4.9	5.0–14.9	15+	Total
Workers employed before 1925					
No. of observed deaths	51	17	13	34	115
No. of expected deaths (adjusted for age and calendar year)	77.58	10.51	10.18	16.73	115.00
Relative risk (using ratios of Observed/Expected)	1.0	2.46	1.94	3.09	
Relative risk (Mantel-Haenszel)	1.0	2.49	2.00	3.14	
Approximate test for homogeneity, $\chi^2_3 = 31.7$; test for trend, $\chi^2_1 = 28.3$ (using observed and expected numbers only with equations (3.11) and (3.12))					
Complete test for homogeneity, $\chi^2_3 = 31.9$; test for trend, $\chi^2_1 = 28.5$ (using full variances with equations (3.24) and (3.25))					
Workers employed in 1925 or after					
No. of observed deaths	100	38	15	8	161
No. of expected deaths (adjusted for age and calendar year)	122.12	22.20	11.04	5.64	161.00
Relative risk (using ratios of Observed/Expected)	1.0	2.09	1.66	1.73	
Relative risk (Mantel-Haenszel)	1.0	2.13	1.64	1.73	
Approximate test for homogeneity, $\chi^2_3 = 17.7$; test for trend, $\chi^2_1 = 10.1$ (using observed and expected numbers only with equations (3.11) and (3.12))					
Complete test for homogeneity, $\chi^2_3 = 17.8$; test for trend, $\chi^2_1 = 10.2$ (using full variances with equations (3.24) and (3.25))					

Table 3.12 Number of men developing nasal sinus cancer by age at first employment and number expected after standardization for year of employment and calendar year of observation[a]

Age at first employment (years)	No. of men developing nasal sinus cancer		Observed as proportion of expected
	Observed	Expected[b]	
Under 20	2	5.36	0.37
20–24	9	11.30	0.80
25–29	13	12.26	1.06
30–34	8	6.34	1.26
35+	8	4.73	1.69
All ages	40	39.99	

χ^2 for trend = 5.2; degrees of freedom (df) = 1; $p = 0.03$

[a] From Doll *et al.* (1970)

[b] If age at first employment had no effect on susceptibility to cancer induction

An example from the literature

A classic example of the use of internal standardization to examine the effect of various time factors on mortality rates is the report of the study of nickel refinery workers in South Wales by Doll *et al.* (1970). The study design is discussed in detail in Appendix ID. Cancer deaths and person-years denominators were classified simultaneously by year of employment (a fixed variable), by age at employment (fixed), and by calendar year of occurrence (time-varying). The effect of each factor was then examined according to the methods described above, using simultaneous stratification on the other two factors. The results shown in Table 3.12 indicate that age at first employment had an influence on the relative incidence of nasal sinus cancer even after the effect of years since exposure (as determined by year of employment and calendar year of observation) had been accounted for. However, calendar year had little effect following adjustment for the other two variables (Table 3.13). The authors concluded: 'The results suggest that, so far as nasal cancer is concerned, susceptibility to induction increases with age and that the risk remains approximately constant for between 15 and 42 years after the carcinogen has been removed from the environment.' The last statement is a reference to the fact that no nasal sinus cancer death was observed among men first employed after 1925, when the manufacturing process was changed. We can agree with these conclusions, provided we bear in mind that they refer to relative risks of cancer mortality rather than absolute ones. Additional analyses of these data which incorporate more recent follow-up are used in Chapters 4, 5 and 6 to illustrate some principles of model fitting.

Table 3.13 Number of men developing nasal sinus cancer by calendar period of observation and number expected after standardization for year and age at first employment[a]

Calendar period of observation	No. of men developing nasal sinus cancer		Observed as proportion of expected
	Observed	Expected[b]	
1939–1941	7	3.63	1.93
1942–1946	8	7.28	1.10
1947–1951	9	9.66	0.93
1952–1956	5	9.34	0.54
1957–1961	6	6.28	0.96
1962–1966	5	3.82	1.31
All years	40	40.01	

χ^2 for trend = 0.95; degrees of freedom (df) = 1; $0.3 < p < 0.5$

[a] From Doll *et al.* (1970)

[b] If year of observation had no effect on risk of developing cancer

3.6 Preferred methods of analysis of grouped data

We repeatedly emphasized in Volume 1 that the goal of a case-control study conducted in a given population was to obtain the same estimates of relative risk as would have been found in a cohort study of that population, had one been performed. Furthermore, methods of analysis of case-control studies were virtually identical to those of cohort studies *vis-à-vis* estimation and testing of hypotheses about relative risk. Thus, it should come as no surprise that the preferred methods of cohort analysis, which we now describe, are nearly identical to those presented in the earlier volume.

The correspondence between case-control and cohort data is easily seen by

comparing the data layout of Table 3.4 with that shown in equation (4.40) of Volume 1. There, we considered the joint distribution of cases (a_{ki}) and controls (c_{ki}) in K exposure groups and I strata; here we deal with deaths and person-years cross-classified into K exposure groups and J strata. Making the substitution of j for i, denoting the cases (deaths) by d_{jk} rather than a_{ki} and considering a fixed number n_{jk} of person-years rather than a random number c_{ki} of controls in each cell, the formal identity of the two situations is complete. All of the test and estimates derived in Volume 1 for a dose-response analysis of case-control data have analogues for use with cohort data. Moreover, the calculations required for cohort data are in most respects even simpler than those for case-control data.

Consider the methods of estimating the relative risk associated with the kth exposure level. For both cohort and case-control studies, these parameters represent the rate ratios for the kth level relative to the first level – ratios that are assumed to remain constant across the various strata. For case-control studies, the odds ratios $(a_{ki}c_{1i})/(a_{1i}c_{ki})$ are good estimates of the corresponding stratum-specific relative risks, and, hence, the analysis may be carried out in terms of summary estimates and tests for heterogeneity and trend in the odds ratios (§2.8, Volume 1). Precisely the same is true of cohort studies, except that the 'odds ratios' $(d_{jk}n_{j1})/(d_{j1}n_{jk})$, rather than being mere approximations to the desired rate ratios, are in fact best estimates of those ratios for the indicated stratum and exposure level.

Some differences between the test statistics used for case-control and cohort studies arise from the different sampling schemes that generate the basic data. In cohort analyses, we regard the observed deaths d_{jk} as having Poisson distributions with means $\psi_k\lambda_{j1}n_{jk}$, where λ_{j1} denotes the baseline death rate in stratum j, and ψ_k is the relative risk associated with exposure at level k. (A more complete statement of this model, its rationale, and its consequences is presented in the next chapter.) If follows that the conditional distribution of the deaths $(d_{j1}, \ldots, d_{jK})$ in each stratum is multinomial with denominator D_j and cell occupancy probabilities $\pi_{jk} = \psi_k n_{jk}/\sum_l \psi_l n_{jl}$). For the case-control study, the conditional distribution of the cases $(a_{1i}, \ldots, a_{Ki})$, given the marginal totals in the $2 \times K$ table (equation 4.40 in Volume 1), was multidimensional hypergeometric with noncentrality parameter depending on the relative risk ψ_k. Differences between the variances of the multinomial and hypergeometric distributions lead to slight differences in the corresponding test statistics. The cohort statistics are simpler because one does not need to consider the marginal totals $d_{jk} + n_{jk}$ at all. By substituting n_{jk} for both c_{ki} and m_i, N_j for both n_{0i} and N_i and d_{jk} for a_{ki}, many of the statistics developed in §4.5 of Volume 1 are converted into precisely the form needed for cohort analyses. Furthermore, just as the tests presented there were derived as efficient score tests based on linear logistic models for binomially distributed case-control data, the versions of those same tests presented here are derived as efficient score tests for analogous hypotheses based on log-linear models for Poisson distributed cohort data.

(*a*) *Two dose levels*: *exposed* versus *unexposed*

Let us start by considering once again the simple problem of comparing death rates for exposed *versus* unexposed without any stratification. We regard O_1 and O_2 as

Poisson variables with means λN_1 and $\psi\lambda N_2$, respectively, where λ represents the background rate, ψ the relative risk, and N_1 and N_2 are the corresponding person-years. Conditional on the total $O_+ = O_1 + O_2$, O_2 is binomially distributed with parameters O_+ and $\pi = \psi N_2/(N_1 + \psi N_2)$. The situation is formally identical to that already considered in §3.4; E_1^* and E_2^* have simply been replaced by N_1 and N_2. Hence, one may apply the same procedures for exact and approximate inferences about π using the binomial distribution and its normal approximation. These are the analogues for cohort analysis of the exact and approximate methods for case-control data developed in §§4.2 and 4.3 of Volume 1.

Example 3.10

Suppose that $O_1 = 5$ lung cancer deaths are observed among a cohort of unexposed persons with $N_1 = 7300$ person-years of observation, whereas $O_2 = 14$ such deaths occur among the exposed with $N_2 = 5500$ person-years of observation. These are precisely the numbers of deaths considered in Example 3.4, and the person-years $N_1 : N_2$ and expected numbers $E_1^* : E_2^*$ are likewise in equal proportion. Consequently, the calculations made earlier apply here as well: $\hat{\psi} = 3.72$ with exact 95% limits of (1.26, 13.2) and approximate ones of (1.25, 11.8).

In more realistic situations, the deaths and person-years are stratified into a series of J 2×2 tables ($j = 1, \ldots, J$) representing different age strata, as shown in Table 3.4. Conditional on fixed values for the total D_j of deaths in the jth stratum, the number of these that occur at the second exposure level is binomially distributed with parameters D_j and $\pi_j = \psi n_{j2}/(n_{j1} + \psi n_{j2})$. Exact inferences about ψ could, in principle, be made from the convolution of these J binomial distributions in the same fashion that exact inferences about the odds ratio in case-control studies are made from the convolution of the corresponding hypergeometric distributions (Gart, 1971). However, the usual normal approximations are entirely satisfactory for most practical purposes.

(*b*) *Summary test of significance*

A test of the null hypothesis $\psi = 1$ is obtained by referring the standardized deviate

$$\chi = \frac{|O_2 - E(O_2)| - 1/2}{\{\mathrm{Var}\,(O_2)\}^{1/2}} = \frac{|O_2 - \sum_{j=1}^{J} n_{j2}D_j/N_j| - 1/2}{\{\sum_{j=1}^{J} D_j n_{j1} n_{j2}/N_j^2\}^{1/2}} \tag{3.14}$$

to tables of the normal distribution. When squared, this is the analogue of the summary statistic used to test for a relative risk of unity in case-control studies (equation 4.23 in Volume 1). Note the use of the continuity correction to improve the normal approximation.

(*c*) *The maximum likelihood estimate*

In large samples the most accurate estimator of ψ is the maximum likelihood estimate, obtained by setting the observed number of deaths O_2 equal to its expected value

$$O_2 = E(O_2; \psi) = \sum_{j=1}^{J} D_j \psi n_{j2}/(n_{j1} + \psi n_{j2}). \tag{3.15}$$

Since solution of (3.15) requires iterative calculations, its use is generally restricted to computer analyses and in particular those which involve the fitting of log-linear models. Note that the problems with maximum likelihood estimation of the common odds ratio in a large series of small 2×2 tables (Breslow, 1981) do not apply to the present situation. Under the Poisson model, conditional and unconditional maximum likelihood estimators are identical (Haberman, 1974).

(*d*) *The Mantel–Haenszel estimate and its standard error*

The Mantel–Haenszel estimate for cohort data is a simple and robust alternative to maximum likelihood. It is written

$$\hat{\psi}_{\mathrm{MH}} = \frac{\sum_{j=1}^{J} R_j}{\sum_{j=1}^{J} S_j} = \frac{\sum_{j=1}^{J} d_{j2} n_{j1}/N_j}{\sum_{j=1}^{J} d_{j1} n_{j2}/N_j}, \tag{3.16}$$

where R_j and S_j are defined by the numerator and denominator terms on the right-hand side of the equation. Clayton (1982) has shown that this estimate arises at the first stage of iteration of one of the computational methods used to find the maximum likelihood estimate. Numerical examples presented below indicate a very good agreement between the two.

A robust variance formula for the Mantel–Haenszel estimate was lacking at the time Volume 1 was written, but the situation has since been remedied both for cohort (Breslow, 1984b) and case-control studies (Robins *et al.*, 1986b). Because of the skewness of the distribution of $\hat{\psi}_{\mathrm{MH}}$ it is more appropriately applied on the log scale. Using the fact that $\hat{\psi}_{\mathrm{MH}} - \psi = \sum_j (R_j - \psi S_j)/\sum_j S_j$, we have the asymptotic

$$\mathrm{Var}\,(\hat{\psi}_{\mathrm{MH}}) = \frac{\sum_{j=1}^{J} \mathrm{Var}\,(R_j - \psi S_j)}{\{\sum_{j=1}^{J} E(S_j)\}^2},$$

and thus that the estimated variance of $\hat{\beta}_{\mathrm{MH}} = \log(\hat{\psi}_{\mathrm{MH}})$ of the log relative risk parameter $\beta = \log(\psi)$ is

$$\mathrm{Var}\,(\hat{\beta}_{\mathrm{MH}}) = \hat{\psi}_{\mathrm{MH}}^{-2}\mathrm{Var}\,(\hat{\psi}_{\mathrm{MH}}) = \frac{\sum_{j=1}^{J} n_{j1} n_{j2} D_j / N_j^2}{\hat{\psi}_{\mathrm{MH}}\left\{\sum_{j=1}^{J} \dfrac{n_{j1} n_{j2} D_j}{N_j(n_{j1} + \hat{\psi}_{\mathrm{MH}} n_{j2})}\right\}^2}. \tag{3.17}$$

Equations (3.16) and (3.17) are symmetric in the sense that interchanging the role of exposed and unexposed subcohorts has the effect of transforming $\hat{\psi}_{\mathrm{MH}}$ into $1/\hat{\psi}_{\mathrm{MH}}$ and $\hat{\beta}_{\mathrm{MH}}$ into $-\hat{\beta}_{\mathrm{MH}}$, but leaves the estimate of $\mathrm{Var}\,(\hat{\beta}_{\mathrm{MH}}) = \mathrm{Var}\,(-\hat{\beta}_{\mathrm{MH}})$ unchanged. Equation (3.17) applies only to Poisson distributed data as collected in a cohort study. The recommended Mantel–Haenszel variance estimate for case-control studies (Robins *et al.*, 1986b) is more complicated.

One important use of any variance estimate is to set approximate confidence intervals on the estimated parameter. Using the interval $\hat{\beta}_{\mathrm{MH}} \pm Z_{\alpha/2}\{\mathrm{Var}\,(\hat{\beta}_{\mathrm{MH}})\}^{1/2}$ for β, we have

$$\psi_L = \hat{\psi}_{\mathrm{MH}} \exp\{-Z_{\alpha/2}(\mathrm{Var}\,\hat{\beta}_{\mathrm{MH}})^{1/2}\}$$

and

$$\psi_U = \hat{\psi}_{\mathrm{MH}} \exp\{+Z_{\alpha/2}(\mathrm{Var}\,\hat{\beta}_{\mathrm{MH}})^{1/2}\}, \tag{3.18}$$

where $\operatorname{Var}\hat{\beta}_{MH}$ is given by (3.17). Alternatively, we could solve iteratively the equations

$$O_2 - 1/2 - \sum_{j=1}^{J} \frac{\psi_L D_j n_{j2}}{(n_{j1} + \psi_L n_{j2})} = +Z_{\alpha/2}\left[\sum_{j=1}^{J} \frac{\psi_L D_j n_{j1} n_{j2}}{(n_{j1} + \psi_L n_{j2})^2}\right]^{1/2}$$

and (3.19)

$$O_2 + 1/2 - \sum_{j=1}^{J} \frac{\psi_U D_j n_{j2}}{(n_{j1} + \psi_U n_{j2})} = -Z_{\alpha/2}\left[\sum_{j=1}^{J} \frac{\psi_U D_j n_{j1} n_{j2}}{(n_{j1} + \psi_U n_{j2})^2}\right]^{1/2},$$

which are based on the notion that $\{O_2 - E(O_2; \psi\}/\{\operatorname{Var}(O_2; \psi)\}^{1/2}$ has an approximate unit normal distribution. These equations are the analogues of equation (4.27) in Volume 1.

Example 3.11

In order to estimate the relative risk associated with 15 or more years moderate or heavy exposure to arsenic among men first employed prior to 1925 in the Montana study, we abstracted 13 2×2 tables from Appendix V giving deaths and person-years for exposure levels 1 and 4. These are shown in Table 3.14. While for most ages, rates are higher in the heavily exposed group, the effect is concentrated particularly in the earlier calendar periods among men aged 50–69. Overall, there are 34 deaths in the higher exposure category, whereas 15.36 would be expected under the null hypothesis that the death rates for the two exposure levels were equal within each of the 13 strata. Since the null variance is 12.42, the summary test statistic (3.14) is $\chi = (34 - 15.36)/\sqrt{12.42} = 5.29$ ($p < 0.0001$). The estimate $\hat{\beta}_{MH} = \log(\hat{\psi}_{MH})$ is $1.144 = \log(3.138)$ and has a standard error calculated according to (3.17) of $\sqrt{\operatorname{Var}(\hat{\beta}_{MH})} = 0.2239$. These values are quite close to those of the maximum likelihood estimate (MLE) $\hat{\beta}_{ML} = 1.126$ and its standard error $\mathrm{SE}(\hat{\beta}_{ML}) = 0.2238$ that were obtained as a by-product of fitting the corresponding model. Approximate confidence limits based on (3.18) are (2.02, 4.87), while those obtained by solving equations (3.19) are (1.94, 4.65). Mantel–Haenszel estimates of relative risk for each of the other exposure categories are shown in Table 3.11.

(e) *Testing for heterogeneity of relative risk (effect modification)*

A fundamental assumption underlying the use of the Mantel–Haenszel or other estimators of relative risk is that the ratio of disease rates between the two exposure categories is constant over the various age groups, calendar years, or other groupings used for stratification of the sample. If there are substantial discrepancies or trends in the disease rate ratios, use of a summary relative risk measure is generally not advisable. Instead, one wants to describe how the effects of exposure as measured by relative risk are modified by age or year. Simple test statistics are available to evaluate this assumption by comparing the observed numbers of deaths among the exposed and unexposed in each stratum with expected numbers calculated using the summary estimate of relative risk. These are closely related to the statistics developed to test for differences between the odds ratios in a series of 2×2 tables formed from case-control data (equations 4.30 and 4.31 in Volume 1).

Setting $\hat{\pi}_j = \hat{\psi} n_{j2}/(n_{j1} + \hat{\psi} n_{j2})$, we denote by $\hat{d}_{j2} = D_j \hat{\pi}_j$, the expected or fitted number of deaths among the exposed and by $\hat{d}_{j1} = D_j(1 - \hat{\pi}_j)$, the number among the unexposed. The maximum likelihood estimator should be used in these calculations, in which case the total number of exposed deaths and the total fitted numbers will agree (equation (3.15)). However, the MH estimator is often sufficiently close to the MLE

Table 3.14 Series of 2 × 2 tables used in example 3.11. Low exposure (−) means less than 1 year of heavy or moderate arsenic exposure; high exposure (+) means 15+ years

Age (years)		Calendar period							
		1938–1949		1950–1959		1960–1969		1970–1977	
	Exposure	−	+	−	+				
40–49	$d/\hat{d}$	2/1.50	0/0.50	0/0.00	0/0.00				
	n	3075.27	337.29	936.75	121.00				
	$\hat{\psi}$	0.00		—					
	Exposure	−	+	−	+	−	+		
50–59	$d/\hat{d}$	2/3.58	4/2.42	3/4.02	3/1.98	3/2.52	1/1.48		
	n	2849.76	626.72	2195.59	349.53	747.77	142.33		
	$\hat{\psi}$	9.0		6.3		1.8			
	Exposure	−	+	−	+	−	+	−	+
60–69	$d/\hat{d}$	2/5.52	9/5.48	7/7.73	7/6.27	10/8.65	3/4.35	1/1.17	1/0.83
	n	2085.43	672.09	1675.91	441.10	1501.73	244.82	440.21	100.64
	$\hat{\psi}$	14.0		3.8		1.8		4.4	
	Exposure	−	+	−	+	−	+	−	+
70–79	$d/\hat{d}$	3/1.98	1/2.02	6/4.32	2/3.68	6/4.40	1/2.60	6/5.62	2/2.38
	n	833.61	277.25	973.32	268.27	1027.12	197.20	674.44	92.75
	$\hat{\psi}$	1.0		1.2		0.9		2.4	

d = observed deaths; $\hat{d}$ = fitted deaths under ML estimate of common rate ratio; n = person-years denominator; $\hat{\psi}$ = rate ratio in each table

that fitted values based on it yield nearly identical results. Moreover, if $O_2 = \sum_j d_{j2}$ and $\sum_j \hat{d}_{j2}$ based on MH differ, say by more than 1%, a 'one-step' correction of $\hat{\beta}_{\text{MH}}$ towards the MLE is available as

$$\hat{\beta}_C = \hat{\beta}_{\text{MH}} + \frac{\sum_{j=1}^{J} d_{j2} - \sum_{j=1}^{J} \hat{d}_{j2}}{\sum_{j=1}^{J} \hat{d}_{j1}\hat{d}_{j2}/D_j}. \tag{3.20}$$

Fitted values $\hat{d}_{j1}$ and $\hat{d}_{j2}$ determined from the corrected MH estimator $\hat{\psi}_C = \exp(\hat{\beta}_C)$ should be adequate for use in what follows if the MLE itself is not available.

To test for a general difference among the rate ratios in the J strata, we compare the

observed and fitted values using the standard chi-square statistic

$$\chi^2_{J-1} = \sum_{j=1}^{J} \left\{ \frac{(d_{j1} - \hat{d}_{j1})^2}{\hat{d}_{j1}} + \frac{(d_{j2} - \hat{d}_{j2})^2}{\hat{d}_{j2}} \right\}, \tag{3.21}$$

which has $J-1$ degrees of freedom. A test for a trend in the stratum-specific ratios with quantitative variables z_j, representing, for example, the age level in stratum j, is accomplished using the statistic

$$\frac{\{\sum_j z_j(d_{j2} - \hat{d}_{j2})\}^2}{\sum_j z_j^2 \hat{d}_{j1}\hat{d}_{j2}/D_j - (\sum_j z_j \hat{d}_{j1}\hat{d}_{j2}/D_j)^2/(\sum_j \hat{d}_{j1}\hat{d}_{j2}/D_j)}. \tag{3.22}$$

This is referred to tables of chi-square on one degree of freedom. If the z_j are equally spaced, the numerator may be reduced in absolute value before squaring by half the distance between adjacent z values in order to correct for the discontinuity of the actual distribution.

Example 3.11 (cont)

Table 3.14 also presents fitted values of $\hat{d}_{j1}$ and $\hat{d}_{j2}$ for the respiratory cancer deaths determined by inserting the MLE $\hat{\psi}_{ML} = \exp(1.126) = 3.083$ in the expressions $\hat{d}_{j1} = D_j n_{1j}/(n_{1j} + \hat{\psi}_{ML} n_{2j})$ and $\hat{d}_{j2} = D_j \hat{\psi}_{ML} n_{2j}/(n_{1j} + \hat{\psi}_{ML} n_{2j})$, respectively. The summary chi-square statistic (3.21) comparing observed and fitted values yields $\chi^2_{12} = 12.9$ ($p = 0.37$), with the largest contribution

$$\frac{(9 - 5.48)^2}{5.48} = 2.25$$

coming from the 60–69-year age group in calendar period 1938–1949. Thus, in spite of the wide range of rate ratios for individual strata observed in this example, the variation is well within the limits expected under the hypothesis that the true ratio is constant across strata.

Example 3.12

Table 3.15 presents data on coronary deaths from the British doctors study (Doll & Hill, 1966) that have been used by Rothman and Boice (1979) and Breslow (1984b) to illustrate methods of cohort analysis. From (3.16) we find a summary relative risk estimate of $\hat{\psi}_{MH} = 1.4247$. The fitted frequencies determined from it

Table 3.15 Deaths from coronary disease among British male doctors[a]

Age group (years)	No. of person-years		No. of observed deaths		No. of expected deaths[b]		Rate ratio	Rate difference per 100 000 person-years
	Non-smokers	Smokers	Non-smokers	Smokers	Non-smokers	Smokers		
j	n_{j1}	n_{j2}	d_{j1}	d_{j2}	$\hat{d}_{j1}$	$\hat{d}_{j2}$		
35–44	18 790	52 407	2	32	6.83	27.17	5.73	50.4
45–54	10 673	43 248	12	104	17.12	98.88	2.14	128.0
55–64	5 710	28 612	28	206	28.74	205.26	1.47	229.6
65–74	2 585	12 663	28	186	26.81	187.19	1.36	385.7
75–84	1 462	5 317	31	102	21.51	111.49	0.90	−202.0
Totals	39 220	142 247	101	630	101.00	630.00	1.72	185.4

[a] Data from Doll and Hill (1966) as quoted by Rothman and Boice (1979)
[b] Estimated by maximum likelihood under the hypothesis of a common rate ratio

total $\sum_j \hat{d}_{2j} = 629.9487$, which agrees very closely with the observed total $\sum_j d_{j2} = 630$. Thus, we know that the MH and MLE estimates are already almost equal, and a correction to the MH estimate would not normally be needed in such circumstances. Nevertheless, in order to illustrate the use of equation (3.20), we further calculate (using fitted values based on $\hat{\psi}_{\text{MH}}$) $\sum_j \hat{d}_{1j}\hat{d}_{j2}/D_j = 88.7729$ and thus find

$$\hat{\beta}_C = 0.35395 + \frac{630 - 629.9487}{88.7729} = 0.35454,$$

which agrees with $\hat{\beta}_{\text{ML}}$ to the number of decimal places shown.

The sixth and seventh columns of Table 3.15 show the final fitted values $\hat{d}_{j1}$ and $\hat{d}_{j2}$ based on $\hat{\psi}_{\text{ML}} = \exp(0.3545) = 1.426$. Inserting these in equations (3.21) and (3.22), and using $z_j = j$ to examine the trend with age, we obtain heterogeneity and (corrected) trend statistics of $\chi^2_4 = 11.1$ and $\chi^2_1 = 10.0$ on four and one degrees of freedom, respectively. Thus, most of the heterogeneity in relative risk is explained by the linear decrease with age. Rothman and Boice (1979) note that these data are more consistent with an additive effect model than with a multiplicative one. In §4.4 we show that an even better fit is obtained using a square-root function to relate age and smoking effects.

In this example, the standard error of $\hat{\beta}_{\text{MH}}$ estimated from the square root of (3.17) is 0.1073, almost identical with $\text{SE}(\hat{\beta}_{\text{ML}}) = 0.1074$. This illustrates once again the generally high efficiency of the MH estimator. However, in other applications the discrepancy may be found to be greater.

(*f*) *Extensions to $K > 2$ exposure classes*

In §3.4, we described the use of externally standardized mortality ratios to evaluate the relative risk of disease associated with each of K exposure categories, for example, the $K = 4$ levels of duration of heavy/moderate exposure to arsenic in both cohorts of the Montana study. A similar approach is taken with the methods of this section. First, Mantel–Haenszel estimates are computed for each of the $k = 2, \ldots, K$ exposure categories relative to baseline. These may be tested for significance individually using the summary chi-square (3.14). The stability of the relative risk estimates from one stratum to another is evaluated using the methods just presented.

In order to test the global null hypothesis that death rates for none of the K exposure classes differ, we require a multivariate extension of (3.14). As in §4.5 of Volume 1, this follows from consideration of the joint distribution, under the null hypothesis, of the deaths $\mathbf{d}_j = (d_{j1}, \ldots, d_{jK})$ in each stratum. Using the Poisson sampling model, the null distribution of $\mathbf{d}_j$ conditional on the total number of deaths D_j in stratum j is easily shown to be multinomial, with a covariance matrix the (k, ℓ) element of which $(1 \leq k, \ell \leq K)$ is

$$\|V_j\|_{k\ell} = \begin{cases} n_{jk}(N_j - n_{jk})D_j/N_j^2, & \text{if } k = \ell \\ -n_{jk}n_{j\ell}D_j/N_j^2, & \text{if } k \neq \ell. \end{cases} \tag{3.23}$$

Under the null hypothesis, the summary vector $\mathbf{O} = (O_1, \ldots, O_K)$ has expectation $\mathbf{E} = (E_1, \ldots, E_K)$ (see equation 3.13) and covariance matrix $\mathbf{V} = \sum_j V_j$. The global test for equality of death rates compares $\mathbf{O}$ and $\mathbf{E}$ using the criterion

$$\chi^2_{K-1} = (\mathbf{O} - \mathbf{E})^{\mathsf{T}}\mathbf{V}^{-}(\mathbf{O} - \mathbf{E}), \tag{3.24}$$

where $\mathbf{V}^{-}$ denotes a generalized inverse of $\mathbf{V}$ (Rao, 1965) and $^{\mathsf{T}}$ denotes a matrix transpose. In practice, this is calculated by restricting $\mathbf{O}$ and $\mathbf{E}$ to the first $K - 1$ components and replacing $\mathbf{V}$ by the corresponding $(K-1) \times (K-1)$ dimensional covariance matrix. For the special case $K = 2$, the test is obtained either as

$(O_1 - E_1)^2/V_1$ or $(O_2 - E_2)^2/V_2$ where $V_1 = \text{Var}(O_1)$ and $V_2 = \text{Var}(O_2)$. It thus reduces to the square of the two-group statistic (3.14), without correction for continuity.

The K group statistic based on indirect standardization, namely (3.11) with the internal fitted values E_k replacing the adjusted external expected, does not require calculation of the variances. It may be recognized as the analogue of the conservative test (equation 4.42 in Volume 1) proposed for case-control data. It always yields smaller values than (3.24), and the degree of conservatism depends on the extent to which the stratum variables confound the disease-exposure relationship (Armitage, 1966; Peto, R. & Pike, 1973).

The test for a trend in relative risk with increasing exposure is obtained from the regression of the observed − expected differences on the dose levels x, namely $\sum_k x_k(O_k - E_k)$. This has a variance of $(\mathbf{x}^T\mathbf{V}\mathbf{x})$. The test statistic is written

$$\chi_1^2 = \frac{\{\sum_{k=1}^K x_k(O_k - E_k)\}^2}{\sum_{k=1}^K x_k^2 E_k - \sum_{j=1}^J (\sum_{k=1}^K x_k e_{jk})^2/D_j}, \tag{3.25}$$

where $e_{jk} = n_{jk}D_j/N_j$ denotes the expected value in the component $2 \times K$ table. This is the analogue of equation (4.43) in Volume 1 for cohort data. Once again, the corresponding statistic based on internal standardization (equation (3.12) using E_k) provides a conservative approximation.

Example 3.13

The last lines in each part of Table 3.11 show the values of the heterogeneity and trend statistics (3.24) and (3.25) obtained with the data in Appendix V. These are slightly greater than the approximating statistics (3.11) and (3.12) calculated from the observed and expected values only, without consideration of the variances. While this is not atypical of what one observes in practice, more serious discrepancies must be anticipated when there is strong confounding.

(g) *Conservatism of indirect standardization*

The impression one might get from our analysis of the Montana smelter workers data, namely, that indirect standardization always yields results close to those obtained with the Mantel–Haenszel methodology, is of course mistaken. The degree of conservatism depends on the degree of statistical confounding between the stratum variables and the exposures. In situations in which the confounding is marked, the conservatism may be also, as the following hypothetical data make clear.

Consider two strata in which the relative risk of exposure is 2, but the pooled risk is considerably less:

	Stratum I		Stratum II		Combined sample	
	Unexposed	Exposed	Unexposed	Exposed	Unexposed	Exposed
Cases	25	5	5	25	30	30
Person-years	10 000	1 000	4 000	10 000	14 000	11 000
Relative risk ($\hat{\psi}$)	2.0		2.0		1.27	
$E_2 = E(O_2)$	27.273		8.571		35.844	
$E_1 = E(O_1)$	2.727		21.429		24.156	
$V_2 = \text{Var}(O_2)$	2.479		6.122		8.602	

The chi-square test for $\psi = 1$ based on internal standardization calculated without continuity correction is

$$\chi^2 = \frac{(30 - 24.156)^2}{24.156} + \frac{(30 - 35.844)^2}{35.844} = 2.37, \quad (p = 0.13),$$

whereas the chi-square test that uses the actual variances from each component table is

$$\chi_1^2 = \frac{(30 - 24.156)^2}{8.602} = 3.97, \quad (p = 0.047).$$

Due to the moderately strong confounding, the approximate statistic is substantially smaller and yields a nonsignificant result. Similarly, the relative risk estimate based only on observed and expected values, namely $(O_2E_1)/(O_1E_2) = (30 \times 35.844)/(30 \times 24.156) = 1.48$, is less than that of the estimate $\hat{\psi}_{MH} = 2.0$.

3.7 Proportional mortality and dose-response analyses

Occasionally one is called upon to conduct a dose-response analysis using only the deaths observed in a defined cohort, without consideration of the corresponding person-years denominators. These may be the only data available. Or, complete exposure histories may have been reconstructed first for dead subjects, for example, and one wants to make an initial evaluation of the probable magnitude of the relative risks before proceeding with the collection of data on those persons who are still alive. The available information consists only of numbers of deaths classified by age at death and other stratification factors, by level of exposure, and by cause of death. Once again, we denote by d_{jk} the number of deaths in stratum j and exposure group k for the cause of interest, by t_{jk} the total deaths from all causes in that stratum and exposure category, and by $D_j = \sum_k d_{jk}$ and $T_j = \sum_k t_{jk}$ the subtotals cumulated over categories. We may also have available a quantitative variable x giving the dose level x_k in exposure class k.

The object of the analysis is to determine whether the proportion of deaths due to the cause of interest increases systematically with increasing levels of exposure, while adjusting for age and other potentially confounding factors by stratification into J strata. The major weakness of the approach is the fact that some of the other causes of death may also be affected by the exposure, thus obscuring the association of interest and hindering precise quantitative estimation of its magnitude. If one is reasonably confident that the other causes of death included in the analysis are not related to the exposure, at least not after accounting for the stratification factors, then the data are best viewed as arising from a type of case-control study in which the deaths from other causes are assumed to represent an unbiased sample (*vis-à-vis* the exposures) of the population at risk within each stratum. This means that the most appropriate analysis of proportional mortality data is to treat them as arising from a case-control study in which the controls died from other causes (Miettinen & Wang, 1981).

In practice, it is useful to exclude from the control sample deaths from those causes that are already known to be related to the exposures. This enhances confidence in the critical assumption that underlies the methodology, namely that the 'controls' are

representative of the population at risk. If one is uncertain about its validity – and this is usually the case – the inferences drawn must necessarily be more tentative than those from an actual case-control study of incident cases in which random sampling methods have been used to select controls from the population in an unbiased fashion.

Although case-control methodology is preferred for the analysis of proportional mortality data, it has been common practice in the past to apply techniques of indirect standardization analogous to those presented in §3.4 and 3.5. One first computes expected numbers of deaths e_{jk} from the cause of interest in the (j, k) stratum/exposure cell under the hypothesis that exposure has had no effect on the death rates. In symbols,

$$e_{jk} = t_{jk} D_j / T_j. \tag{3.26}$$

These values are cumulated to give $E_k = \sum_j e_{jk}$ as the total number expected at level k after adjustment. It would be tempting to insert such E_k into equations (3.8), (3.11) and (3.12) in order to estimate and make tests on the relative risk. If the disease is common, however, such ad-hoc methods may lead to results that are at considerable variance from those obtained using the proper case-control methods. The main difficulty is the fact that the disease of interest is making a contribution to the totals t_{jk} and T_j used to calculate the expected numbers, so that these are closer to the observed numbers than they are for the analogous cohort data. Even under the proportionality assumption that justified dose-response analysis of SMR_k, the equivalent proportional mortality analysis may not be valid.

Example 3.14

The sixth column of Appendix V shows the total numbers of deaths among the Montana workers classified by age, calendar period, date of employment and exposure duration. These were used in a case-control dose-response analysis according to the methods presented in Chapter 4 of Volume 1. There were 18 age × calendar period strata and four exposure levels for the pre-1925 cohort, and 16 strata and four exposure levels for the post-1925 cohort. Table 3.16 presents the results. The Mantel–Haenszel estimates of relative risk are in reasonable agreement with those found from the entire set of cohort data (Table 3.11), except for the highest exposure duration category in the early cohort (2.62 *versus* 3.14). Here, the proportional mortality analysis yields a substantially lower estimate of relative risk, suggesting that causes of death other than respiratory cancer may be affected by lengthy exposures to arsenic. The Mantel–Haenszel estimates are in good agreement with those obtained by (unconditional) maximum likelihood according to the methods presented in Chapter 6 of Volume 1, namely, the fitting of linear logistic models to the binomial proportions of cause-specific deaths divided by total deaths. The statistics (4.41) and (4.43) in Volume 1 for testing for heterogeneity and trend in the relative risks are substantially less than the corresponding statistics shown in the sixth row of Table 3.11 for the full cohort data. This is not surprising in view of the reduced value for the relative risk estimate for the highest exposure category.

Also shown in Table 3.16 for each subcohort are the expected numbers of respiratory deaths obtained by multiplying the total deaths in each age-stratum-exposure cell by the proportion of respiratory deaths in that stratum as shown in equation (3.26), and then summing across strata. When these are inserted in equation (3.8) to estimate the 'relative risk' for each exposure level, the results are considerably more conservative than were the results based on indirect standardization using the complete set of cohort data (Table 3.11). For example, the estimate of relative risk from proportional mortality data for the 15+ years exposure duration category in the pre-1925 cohort is $\hat{\psi} = 2.38$ based on observed/expected values *versus* $\psi = 2.62$ for Mantel–Haenszel. The corresponding figures from cohort data were 3.09 *versus* 3.14. Similarly, whereas the test statistics (3.11) and (3.12) yielded only slightly conservative results when used with internally standardized expected numbers based on the person-years denominators, when used with the proportional expected values from equation (3.26) that depend only on the proportional mortality data, the results are

Table 3.16 Dose–response analysis of respiratory cancer deaths among Montana smelter workers, based on proportional mortality

	Cumulative years of moderate/heavy arsenic exposure				
	0–0.9	1.0–4.9	5.0–14.9	15+	Total
Workers employed prior to 1925					
Observed deaths	51	17	13	34	115
Total deaths (All causes)	636	100	93	195	1024
Expected deaths (Internal adjustment for age and calendar year)	72.47	11.84	10.41	20.29	115.00
Relative risk (using ratios of Observed/Expected)	1.0	2.04	1.77	2.38	
Relative risk (Mantel-Haenszel)	1.0	2.30	2.12	2.62	
Relative risk (Maximum likelihood)	1.0	2.32	1.98	2.82	
Approximate test for homogeneity, $\chi^2_3 = 18.5$; test for trend, $\chi^2_1 = 16.6$ (using observed and expected values in equations (3.11) and (3.12))					
Case-control test for homogeneity, $\chi^2_3 = 21.6$; test for trend, $\chi^2_1 = 19.4$ (equations 4.41 and 4.43 from Volume 1)					
Workers employed 1925 or later					
Observed deaths	100	38	15	8	161
Total deaths (All causes)	1389	274	143	68	1874
Expected deaths (Internal adjustment for age and calendar year)	118.47	24.47	11.83	6.25	161.00
Relative risk (using ratios of Observed/Expected)	1.0	1.84	1.50	1.52	
Relative risk (Mantel-Haenszel)	1.0	2.06	1.54	1.58	
Relative risk (Maximum likelihood)	1.0	2.02	1.57	1.61	
Approximate test for homogeneity, $\chi^2_3 = 11.7$; test for trend, $\chi^2_1 = 6.3$ (using observed and expected values in equations (3.11) and (3.12))					
Case-control test for homogeneity, $\chi^2_3 = 13.2$; test for trend, $\chi^2_1 = 7.0$ (equations 4.41 and 4.43 from Volume 1)					

noticeably different from those obtained with proper case-control techniques. This illustrates the basic point that indirect standardization techniques should not be used in the context of proportional mortality unless one is dealing with a very rare disease. Whereas they may or may not yield conservative results with cohort data depending on the degree of statistical confounding, they are bound to produce conservative results with proportional mortality (case-control) data.

Nothing has yet been said about the possibility of incorporating information from the external standard population into the dose-response analysis of proportional mortality data. The reason is that the elementary methods presented in §3.4 for cohort studies

have no suitable analogue when death records are the only data available, nor do the indirect standardization techniques of §3.5, as shown in the preceding example. Suppose we were to calculate expected numbers of deaths for each exposure category using the formula $E_k^* = \sum_j t_{kj} p_j^*$, where p_j^* denotes the standard proportion of deaths in stratum j due to the cause of interest. Even under the assumption of proportionality, in which the stratum-specific mortality rates for both cause-specific and general deaths in each exposure category are constant multiples of the stratum-specific standard rates, inserting these expected numbers into equations (3.8), (3.11) and (3.12) may yield badly biased estimates and tests if more than a few percent of total deaths are due to the cause of interest. Although it is possible to use the external standard proportions by incorporating them into an appropriate model, none of the standard estimates or tests based on the model have simple closed form expressions. Therefore, we defer further discussion of this approach to proportional mortality analysis until the next chapter.